



```

Bacillus amyloliquefaciens contain a large open reading
frame between the regions coding for signal sequence and
mature protein.
#cross-references MUID:85006739
#accession A25415
##molecule_type DNA
##residues 1-521 ##label YAN
##cross-references GB:K02497; NID:g143248; PID:g143249
GENETICS
#gene npr
#start_codon GTG
CLASSIFICATION #superfamily thermolysin
KEYWORDS calcium; extracellular protein; hydrolase; metalloproteinase;
zinc
FEATURE
1-27 . domain signal sequence #status predicted #label SIG\
28-221 . domain propeptide #status predicted #label PRO\
222-521 . product bacillolysin #status predicted #label MAT\
364,368,388 . binding_site zinc (His, His, Glu) #status predicted\
365,449 . active_site Glu, His #status predicted
SUMMARY
#length 521 #molecular-weight 56840 #checksum 7195
Query Match 14.5%; Score 92; DB 1; Length 521;
Best Local Similarity 27.8%; Pred. No. 1.99e-01;
Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps 2;
Db 240 ISSSEKGYVLRLSKPTGTQIIITYDLQNRVNLPGTLYSSTNQFTSSORAA 292
QY 36 LSTAGPSYV-KFQDNVPVGSOTFSAGLHLRVDPSTGALVDSKSYAFSTNDTTS 88

RESULT 3
ENTRY G71525 #type complete
TITLE Probable OMP [leader (19) peptide] - Chlamydia trachomatis
(seroType D, strain UW3/Cx)
ORGANISM #formal_name Chlamydia trachomatis
DATE 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change
21-Nov-1998
ACCESSIONS G71525
REFERENCE A71570
#authors Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe,
R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov,
R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
#journal Science (1998) 282:754-759
#title Genome sequence of an obligate intracellular pathogen of
humans: Chlamydia trachomatis.
#cross-references MUID:99000809
#accession G71525
##status preliminary
##molecule_type DNA
##residues 1-566 ##label ARN
##cross-references GB:AE001308; GB:AE001273; NID:g3328766; PID:g3328771
##experimental_source serotype D, strain UW-3/Cx
GENETICS
#gene CT350
SUMMARY
#length 566 #molecular-weight 63507 #checksum 4960
Query Match 14.0%; Score 89; DB 2; Length 566;
Best Local Similarity 26.2%; Pred. No. 5.25e-01;
Matches 17; Conservative 25; Mismatches 18; Indels 5; Gaps 5;
Db 63 IAESYLOQSFSEDTYIR-KSAITGA-GLGSSSEA-LELLSEATETQDYEQL-L-ILNA 117
QY 8 FGQGYVTPELSENSVRYKISITAGSCPLSTAGPSYV-KFQDNVPVGSOTFSAGLHLRVDP 67
Db 118 ATSQL 122
QY 68 STGAL 72

RESULT 4
ENTRY A29091 #type complete
TITLE licheninase (EC 3.2.1.73) beta - Bacillus amyloliquefaciens

```

```

ALTERNATE_NAMES 1.3-1.4-beta-D-glucan 4-glucanohydrolase; beta-glucanase;
lichenase
ORGANISM #formal_name Bacillus amyloliquefaciens
DATE 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change
20-Mar-1998
ACCESSIONS A29091
REFERENCE A91564
#authors Hofemeister, J.; Kurtz, A.; Borriess, R.; Knowles, J.
#journal Gene (1986) 49:177-187
#title The beta-glucanase gene from Bacillus amyloliquefaciens shows
extensive homology with that of Bacillus subtilis.
#cross-references MUID:87192007
#accession A29091
##molecule_type DNA
##residues 1-239 ##label HOF
##cross-references GB:M15674; NID:g143009; PID:g143010
##experimental_source strain BE20/78
GENETICS
#gene bglA
CLASSIFICATION #superfamily licheninase
KEYWORDS glycosidase; hydrolase; polysaccharide degradation
SUMMARY
#length 239 #molecular-weight 26928 #checksum 1611
Query Match 13.6%; Score 86; DB 2; Length 239;
Best Local Similarity 25.5%; Pred. No. 1.35e+00;
Matches 14; Conservative 20; Mismatches 18; Indels 3; Gaps 3;
Db 74 LAUTSPYKFKDCGNSVQTYGYLYEVRM-KPAKNTGIVSSFFTYGTGTEGTP 127
QY 36 LSTAGPSYV-KFQDNVPVGSOTFSAGLHLRVDPSTGALVDSKSYAFSTNDTTS 88

RESULT 5
ENTRY I39956 #type complete
TITLE neutral proteinase (EC 3.4.24.-) - Bacillus amyloliquefaciens
ORGANISM #formal_name Bacillus amyloliquefaciens
DATE 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change
18-Mar-1997
ACCESSIONS I39956
REFERENCE I39956
#authors Shimada, H.; Honjo, M.; Mita, I.; Nakayama, A.; Akaoka, A.;
Manabe, K.; Furutani, Y.
#journal J. Biotechnol. (1985) 2:75-85
#title The nucleotide sequence and some properties of the neutral
protease gene of Bacillus amyloliquefaciens.
#accession I39956
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-521 ##label RES
##cross-references GB:M36723; NID:g143352; PID:g143353
GENETICS
#start_codon GTG
CLASSIFICATION #superfamily thermolysin
KEYWORDS hydrolase; metalloproteinase
SUMMARY
#length 521 #molecular-weight 56725 #checksum 6816
Query Match 13.6%; Score 86; DB 2; Length 521;
Best Local Similarity 27.8%; Pred. No. 1.35e+00;
Matches 15; Conservative 18; Mismatches 19; Indels 2; Gaps 2;
Db 240 ISSSEKGYVLRLSKPTGTQIIITYDLQNRVNLPGTLYSSTNQFTSSORAA 292
QY 36 LSTAGPSYV-KFQDNVPVGSOTFSAGLHLRVDPSTGALVDSKSYAFSTNDTTS 88

RESULT 6
ENTRY S28275 #type fragment
TITLE hypothetical protein F54G8.4 - Caenorhabditis elegans
(fragment)
ORGANISM #formal_name Caenorhabditis elegans
DATE 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change
09-Sep-1997
ACCESSIONS S28275

```

[illegible]

Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denicot, F.; Devine, K.M.; Duesterhoeft, A.J.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, S.; Fujita, Y.; Funa, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golligly, E.J.; Grandi, G.; Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinols, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestil, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Potwollik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemura, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

Nature (1997) 390:249-256

The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

#journal Nature (1997) 390:249-256

#title The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

#cross-references MUID:98044033

#accession B59594

##status preliminary; nucleic acid sequence not shown; translation not shown

##molecule\_type DNA

##residues 1-242 ##label KUN

##cross-references GB:299124; GB:AL009126; NID:g2636442; PID:el184632;

##experimental\_source strain C120

##experimental\_source PID:g2636453

##experimental\_source A93526

##authors Murphy, N.; McConnell, D.J.; Cantwell, B.A.

#journal Nucleic Acids Res. (1984) 12:5355-5367

#title The DNA sequence of the gene and genetic control sites for the excreted *B. subtilis* enzyme beta-glucanase.

#cross-references MUID:8427222

#accession A2914

##molecule\_type DNA

##residues 1-203, 'L', 205-242 ##label MUR

##cross-references EMBL:X00754; NID:g39818; PID:g685236

##experimental\_source strain C120

##note the authors translated the codon CAA for residue 29 as Lys and CCA for residue 82 as Leu

A90026

#authors Tezuka, H.; Yuuki, T.; Yabuuchi, S.

#journal Agric. Biol. Chem. (1989) 53:2335-2339

#title Construction of a beta-glucanase hyperproducing *Bacillus subtilis* using the cloned beta-glucanase gene and a multi-copy plasmid.

#accession A90026

##molecule\_type DNA

##residues 1-23, 'S', 25-82, 'S', 84-242 ##label TEZ

##cross-references DBJ:D00518; NID:g216243; PID:g216244

##experimental\_source strain Y-25, clone pLE100

A90027

#authors Yuuki, T.; Tezuka, H.; Yabuuchi, S.

#journal Agric. Biol. Chem. (1989) 53:2341-2346

#title Purification and some properties of two enzymes from a beta-glucanase hyperproducing strain, *Bacillus subtilis* HL-25







工 口 司 同 向

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.

Result No.	Score	Query		Length	DB	ID	Description	Pred. No.
		Match						
1	300	47.3	65	1	V15A_BP74		HYPOTHETICAL 7.3 KD PR	1.50e-44
2	100	15.8	1693	1	POLN_HEVW		NON-STRUCTURAL POLYPRO	3.02e-03
3	92	14.5	521	1	NPRE_BACAM		BACILLOLYSIN PRECURSOR	5.71e-02
4	87	13.7	703	1	G41I_RHSN		HYPOTHETICAL 76.2 KD P	3.31e-01
5	86	13.6	239	1	YUB_BACAM		BETA-GLUCANASE PRECURS	4.66e-01
6	86	13.6	457	1	ENB8_PICGL		LATE EMBRYOGENESIS ABU	4.66e-01
7	86	13.6	672	1	NOO3_PARDE		NADH-UBIQUINONE OXIDOR	4.66e-01
8	86	13.6	932	1	YMB4_CAEEL		HYPOTHETICAL 104.4 KD	4.66e-01
9	86	13.6	1693	1	POLN_HEYBU		NON-STRUCTURAL POLYPRO	4.66e-01
10	86	13.6	1693	1	POLN_HEVPA		NON-STRUCTURAL POLYPRO	4.66e-01
11	85	13.4	242	1	GUB_BACSU		BETA-GLUCANASE PRECURS	6.55e-01
12	85	13.4	869	1	CFAC_ECOLI		CFA/I FIMBRIAL SUBUNIT	6.55e-01
13	84	13.2	243	1	GUB_BACLI1		BETA-GLUCANASE PRECURS	9.18e-01
14	83	13.1	611	1	YX45_MYCTU		HYPOTHETICAL 67.2 KD P	1.28e+00
15	82	12.9	449	1	YM60_YEAST		HYPOTHETICAL 51.4 KD P	1.79e+00
16	82	12.9	527	1	TXK_HUMAN		TYROSINE-PROTEIN KINAS	1.79e+00
17	81	12.8	345	1	GCH2_YEAST		GTP CYCLOHYDROLASE II	2.48e+00
18	81	12.8	433	1	ENO_BORBU		ENOLASE (EC 4.2.1.11)	2.48e+00
19	81	12.8	527	1	TXK_MOUSE		TYROSINE-PROTEIN KINAS	2.48e+00
20	80	12.6	171	1	YCB2_PSEDE		HYPOTHETICAL 19.0 KD P	3.43e+00
21	80	12.6	300	1	NTRY_AZOBR		NITROGEN REGULATION PR	3.43e+00
22	80	12.6	572	1	CIY42_TRYBB		RECEPTOR-TYPE ADENYLAT	3.43e+00
23	80	12.6	802	1	NAB3_YEAST		NUCLEAR POLYADENYLATED	3.43e+00



This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement or send an email to licensee@isb-sib.ch).

EMBL; L47118; GI350545; -.  
DR PROSITE; PS01133; UPF0017; 1.  
SQ SEQUENCE 457 AA; 51019 MW; AE7CB4CD CRC32;

Query Match 13.6%; Score 86; DB 1; Length 457;  
Best Local Similarity 34.0%; Pred. No. 4.66e-01;  
Matches 17; Conservative 11; Mismatches 20; Indels 2; Gaps 2;

Dd 232 LGANILRYLGCVAGNCPLSGVSLCNPP-NLVIAEDFHKGGLGFNNYD 280  
:: :: ||| :|||:: | : : : : | : | :  
Qy 18 LSSENSVRKIISGIASCPLSTAGPSYYKFDNPVPVSQTFSAGHLRL-VFD 66

RESULT 7

ID NOO3 PARDE STANDARD; PRT; 672 AA.

AC A239915;  
DT 01-APR-1993 (REL. 25, CREATED)  
DC 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE NADH+UBIQUINONE OXIDOREDUCTASE CHAIN 3 (EC 1.6.5.3) (NADH  
DHYDROGENASE 1, CHAIN 3) (NDH-1, CHAIN 3).  
GN NOO3.  
OS PARACOCCUS DENITRIFICANS.  
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHODOBACTER GROUP;  
RC PAROCOCCUS.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.  
RX STRAIN=ATCC 13348;  
RK MEDLINE; 92296779.  
RA XU X., MATSUONO-YAGI A., YAGI T.;  
RT "Structural features of the 66-kDa subunit of the energy-transducing  
RL NADH-ubiquinone oxidoreductase (NDH-1) of Paracoccus denitrificans.";  
RR ARCH. BIOCHEM. BIOPHYS. 296:40-48(1992).  
[2]

RZ SEQUENCE OF 657-672 FROM N.A.  
RX MEDLINE; 93136200.  
RU XU X., MATSUONO-YAGI A., YAGI T.;  
RV "DNA sequencing of the seven remaining structural genes of the gene  
cluster encoding the energy-transducing NADH-quinone oxidoreductase  
of Paracoccus denitrificans";  
RW BIOCHEMISTRY 32:968-981(1993).  
XC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.  
CC -1- COFACTOR: MAY BIND TWO 4FE-4S CLUSTER AND ONE 2FE-2S CLUSTER.  
CC -1- SUBUNIT: COMPOSED OF 14 DIFFERENT SUBUNITS.  
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 75 KD SUBUNIT FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement or send an email to licensee@isb-sib.ch).

EMBL; M84572; G150601; -.  
DR PIR; S23948; S23948.  
DR PROSITE; PS00641; COMPLEX1\_75K\_1; 1.  
DR PROSITE; PS00642; COMPLEX1\_75K\_2; 1.  
DR PROSITE; PS00643; COMPLEX1\_75K\_3; 1.  
DR PFAM; PF00111; fer2; 1.  
DR PFAM; PF00384; molybdopterin; 1.  
KW OXIDOREDUCTASE; NAD; UBIQUINONE; IRON-SULFUR; 4FE-4S.  
FT INIT\_MET 0  
FT METAL 25 25 IRON-SULFUR (2FE-2S) (POTENTIAL).  
FT METAL 36 36 IRON-SULFUR (2FE-2S) (POTENTIAL).

```

Db      282 TFTGLEVLRNFQPDADFLNRK-LRFSSGDDAA 314
||: || || ||: | :| ||::||:
QY      55 TFSAGHLR-LRVFDPSTGALVDKSYAFSTNDTT 87

RESULT          9
ID   POLN.HEVBU    STANDARD;          PRT; 1693 AA.
AC   P29324;
DT   01-DEC-1992 (REL. 24, CREATED)
DI   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE
DS   (EC 2.7.7.48); HELICASE].
OS   HEPATITIS E VIRUS (STRAIN BURMA) (HEV).
OC   VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;
QC   CALICIVIRUS.
RN   [1]
RX   SEQUENCE FROM N.A.
RY   MEDLINE; 92024067.
RA   TAM A.W., SMITH M.M., GUERRA M.E., HUANG C.-C., BRADLEY D.W.,
RA   FRY K.E., REYES G.R.;
RT   "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT   full-length viral genome.";
RL   VIROLOGY 185:120-131(1991).
CC   -1- HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF ENTERICALLY
CC   TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
-----
CC   THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-----
CC   EMBL; M73218; G330024; -
DR   PIR; A40778; MNWHE.
KW   POLYPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; HELICASE;
KT   ATP-BINDING.
FT   NP_BIND     975      982      ATP (POTENTIAL)..
SQ   SEQUENCE 1693 AA; 185191 MW; C560BE14 CRC32;

Query Match      13.6%; Score 86; DB 1; Length 1693;
Best Local Similarity 25.4%; Pred. No. 4.66e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db      572 PRTSFDGAVLENGPRHNLSSTDASQTWAAGFSITYAASAGLEVRYVAAGLDHRV 631
||: ||: | :| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
QY      8 FGOGYVOTPELSRSNVRYKISIAGSCPLSTAGPSYVKFQDNPVGSQT-F-SAGLHLR-V 64

Db      632 FAGGVSP 638
||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
QY      65 FDPSTGA 71

RESULT         10
ID   POLN.HEVPA    STANDARD;          PRT; 1693 AA.
AC   P33424;
DT   01-FEB-1994 (REL. 28, CREATED)
DI   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE
DS   (EC 2.7.7.48); HELICASE].
OS   HEPATITIS E VIRUS (STRAIN PAKISTAN) (HEV).
OC   VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;
QC   CALICIVIRUS.
RN   [1]
RX   SEQUENCE FROM N.A.
RY   MEDLINE; 92115700.
RA   TSAREV S.A., EMERSON S.U., REYES G.R., TSAREVA T.S., LEGTERS L.J.,
RA   MALIK I.A., IQBAL M., PURCELL R.H.;
RT   "Characterization of a prototype strain of hepatitis E virus.";
```



```

QG PLASMID NTP513
OC BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
[1]
RN RN SEQUENCE FROM N.A.
RN RN STRAIN-ENTEROTOXIGENIC;
RX RX MEDLINE: 89330163.
RA HAWES A.M., PEL H.J., WILLISAW G.A., KUSTERS J.G.,
RA VAN DER ZEIJST B.A.M., GAASTRA W.;
RT "The nucleotide sequence of the first two genes of the CFA/I fimbrial
RL operon of human enterotoxigenic Escherichia coli.";
RL MICROB. PATHOG. 6:297-309(1989).
RN RN [2]
RN RN SEQUENCE FROM N.A.
RX RX MEDLINE: 92329981.
RA JORDI B.J.A.M., WILLISAW G.A., VAN DER ZEIJST B.A.M., GAASTRA W.;
RT "The complete nucleotide sequence of region 1 of the CFA/I fimbrial
RL operon of human enterotoxigenic Escherichia coli.";
RL DNA SEQ. 2:257-263(1992).
CC CC -1- FUNCTION: MAY SERVE AS ANCHOR FOR THE FIMBRIAE IN THE OUTER
CC CC MEMBRANE.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DB DR EMBL: M55661; GI45510; -
KW ANTIGEN; SIGNAL; FIMBRIA; OUTER MEMBRANE; PLASMID.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 869 CFA/I FIMBRIAL SUBUNIT C.
SQ SEQUENCE 869 AA: 97830 MW; 7AF76347 CRC32;
Query Match 13.4%; Score 85; DB 1; Length 869;
Best Local Similarity 32.4%; Pred. No. 6.55e-01;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;
Db 158 AF105QTINLSDSGKRYKRLISGNSALGITDTSY 191
:1: :1: :1: :1: :1: :1: :1: :1: :1:
QY 11 GYVQTPFLSESNVRYK-ISIAGSCPLSTAGPSY 43
:1: :1: :1: :1: :1: :1: :1: :1: :1:
RESULT 13
ID GUB_BACLI STANDARD; PRT; 243 AA.
AC P27051;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
DE BG1.
OS BACILLUS LICHENIFORMIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
[1]
RN RN SEQUENCE FROM N.A.
RX RX MEDLINE: 91224124.
RA LLOBERAS J., PEREZ-PONS J.A., QUEROL E.;
RT "Molecular cloning, expression and nucleotide sequence of the
RT endo-beta-1,3-1,4-D-glucanase gene from Bacillus licheniformis.
RT Predictive structural analyses of the encoded polypeptide.";
RL EUR. J. BIOCHEM. 197:337-343(1991).
RN RN [2]
RN RN REVISIONS.
RA QUEROL E.;
RL SUBMITTED (JUL-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
RN RN [3]
RN RN MUTAGENESIS.
RX RX MEDLINE: 92362869.
RA PLANAS A., JUNCOSA M., LLOBERAS J., QUEROL E.;

```



```
Query Match      12.9%; Score 82; DB 1; Length 449;
Best Local Similarity 27.1%; Pred. No. 1.79e+00;
Matches 16; Conservative 18; Mismatches 22; Indels 3; Caps 3;
```

(TM)

Tabular output not generated.

Description: (1-93) from PCTUS9913024.pep (6 of 12)

sequence: I MEKFMAEFGQYVQTPFLSE.....DSKSYAFSTSDTTSAAFVS 93

Gap 11

Searched: 179066 seqs, 54579741 residues

Listing first 45 summaries

12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 37.991; Variance 71.901; scale 0.528

and is derived by analysis of the total score distribution.

## SUMMARIES

No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	101	15.9	1577	2	Q54178	GLUCOSYLTRANSFERASE.	1.06e-02
2	100	15.8	1693	14	Q81876	COMPLETE GENOME SEQUEN	1.50e-02
3	96	15.1	114	14	Q41088	AS60L PROTEIN.	5.83e-02
4	92	14.5	299	14	Q65574	HYPOTHETICAL 32.4 KD P	2.20e-01
5	92	14.5	300	14	Q65574	UL7 POLYPEPTIDE.	2.20e-01
6	92	14.5	337	2	Q44687	NEUTRAL PROTEASE.	2.20e-01
7	89	14.0	566	2	Q84354	NEUTRAL PROTEASE (FRAG	2.20e-01
8	89	14.0	921	5	P90770	HYPOTHETICAL 63.5 KD P	5.81e-01
9	88	13.9	180	10	Q49984	C34B7.2 PROTEIN.	5.81e-01
10	87	13.7	239	2	Q45691	PUTATIVE ETHYLENE REGE	7.99e-01
11	86	13.6	521	2	Q44677	ENDO-BETA-1,3-1,4-GLUC	1.10e+00
12	86	13.6	974	5	Q23510	NEUTRAL PROTEASE.	1.51e+00
13	86	13.6	1693	14	Q69410	F34G8.4 PROTEIN.	1.51e+00
14	86	13.6	1693	14	Q89444	METHYL TRANSFERASE.	1.51e+00
15	86	13.6	1693	14	Q81344	UNNAMED PROTEIN PRODC	1.51e+00
16	86	13.6	1693	14	Q81862	HEPATITIS E VIRUS COM	1.51e+00
17	85	13.4	229	10	P93830	ORF 1.	1.51e+00
18	84	13.2	229	10	Q49462	EARLY AUXIN-INDUCED PR	2.06e+00
19	84	13.2	622	2	Q45743	IAA17/AAK3-1 PROTEIN.	2.81e+00
20	84	13.2	3443	14	Q11979	CYCIC DELTA-ENDOTOXIN	2.81e+00
21	84	13.2	3443	14	Q11979	POLYPROTEIN.	2.81e+00

[illegible]



[illegible]

Search completed: Wed Sep 1 16:16:22 1999  
Job time : 28 secs.

\*\*\*\*\*  
[M][A][P][S][A][R][C][H] (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 15:19:58 1999; MasPar time 7.22 Seconds  
Tabular output not generated. 294.721 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (1-100) from PCTUS9913024.pep (7 of 12)  
Perfect Score: 583  
Sequence: 1 MEKFMAEFGQGVQTPFLSE.....STSDTTSAAFVSPFMNSLTN 100

Scoring Table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 27.671; Variance 122.380; scale 0.226

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	300	43.9	65	19	R97371	Phage T4 ORF gene pr
2	96	14.1	1693	35	W76368	Hepatitis E virus hol
3	92	13.5	520	1	P94617	Neutral protease enco
4	92	13.5	521	3	P51009	Sequence of neutral p
5	89	13.0	438	20	W06355	Malate permease.
6	89	13.0	438	23	W14558	Malate permease.
7	87	12.7	242	19	P95000	Bacillus subtilis lic
8	87	12.7	623	35	W5773	Amino acid sequence o
9	86	12.6	239	1	R06621	Hybrid (1.3-1.4)-pre
10	86	12.6	1693	37	W81519	Hepatitis E virus (HE
11	86	12.6	1693	36	W80196	Protein encoded by OR
12	86	12.6	1693	10	R51264	HEV strain protein en
13	86	12.6	1693	34	W71209	Protein encoded by OR
14	86	12.6	1693	19	R91813	Hepatitis E virus str
15	85	12.4	633	35	W57774	Amino acid sequence o
16	84	12.3	1693	3	R14618	Protein encoded by OR

17	83	12.2	473	21	W14005	Human SHC protein.	4.33e+01
18	83	12.2	473	15	R84637	SHC protein.	4.33e+01
19	83	12.2	474	17	R97243	SHC phosphotyrosine b	4.33e+01
20	83	12.2	613	15	R74632	OETR ethylene respons	4.33e+01
21	83	12.2	613	36	W73122	A. thaliana ethylene	4.33e+01
22	82	12.0	551	24	W18790	Corrected Bacillus la	5.11e+01
23	82	12.0	633	35	W75775	Amino acid sequence o	5.11e+01
24	81	11.9	345	10	R52824	GTP-cyclohydrolase II	6.03e+01
25	81	11.9	3011	8	R40120	HCV genomic amino aci	6.03e+01
26	81	11.9	3011	12	R66995	Hepatitis C virus gen	6.03e+01
27	80	11.7	170	1	P90150	Sequence of hepatitis	7.11e+01
28	80	11.7	170	1	P92033	Sequence encoded in t	7.11e+01
29	80	11.7	411	16	R90934	HCV NS5 domain antige	7.11e+01
30	80	11.7	499	38	W67010	HCV non-structural pr	7.11e+01
31	80	11.7	509	20	W08378	Brassica napus micros	7.11e+01
32	80	11.7	509	2	R06519	Microspore-specific c	7.11e+01
33	80	11.7	516	7	R33633	HCV CFS-NS5E fusion p	7.11e+01
34	80	11.7	516	4	R21566	HCV CFS-NS5E - pHCV-4	7.11e+01
35	80	11.7	1786	1	P92041	Sequence encoded in t	7.11e+01
36	80	11.7	1786	1	P90158	Protein sequence of h	7.11e+01
37	80	11.7	2261	1	P90164	Peptide encoded by co	7.11e+01
38	80	11.7	2301	1	P92047	Sequence encoded in t	7.11e+01
39	80	11.7	2436	5	R28582	HCV amino acid sequen	7.11e+01
40	80	11.7	2482	1	P90288	Peptide encoded by co	7.11e+01
41	80	11.7	2816	7	R34009	HCV-1 polyprotein.	7.11e+01
42	80	11.7	2955	2	R08124	Hepatitis C virus put	7.11e+01
43	80	11.7	3011	28	W40038	HCV polyprotein.	7.11e+01
44	80	11.7	3011	4	R21519	Compiled HCV sequence	7.11e+01
45	80	11.7	3011	26	W34480	HCV polyprotein.	7.11e+01

ALIGNMENTS

RESULT 1  
ID R97371 standard; Protein; 65 AA.  
AC R97371;  
DT 07-JAN-1997 (first entry)  
DE Phage T4 ORF gene product \_gp34.  
KW Phage T4; tail fibre protein; nanotechnology; nano-structure;  
KW filter; molecular sieve.  
OS Bacteriophage T4.  
PN WO9611947-Al.  
PD 25-APR-1996.  
PR 13-OCT-1994; US-322760.  
PF (GOLD/) GOLDBERG E B.  
PI Golberg ES;  
DR WPI; 96-221942/22.  
DR N-PSDB; T29053.  
PT New proteins derived from T4 phage tail fibre proteins - that can self assemble into nano-structure(s), useful as filters etc, also corresponding DNA  
PS Claim 7; Fig 7; 83pp; English.  
CC A protein (R97371) of unspecified function is the product of open reading frame x of the tail fibre protein gene region (see also T29053) of phage T4. This gene region also includes open reading frames for tail fibre proteins (see also R97370 and R97372-74).  
CC Tail fibre proteins (native or modified) can be produced in large quantities in microbial cells and used as building blocks of strong, stable nanostructures.  
SQ Sequence 65 AA;  
  
Query Match 43.9%; Score 300; DB 19; Length 65;  
Best Local Similarity 82.1%; Pred. No. 1.46e+17;  
Matches 46; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Db 1 mekfmaelwtricpnallseesvrykisiagscplstagspsvkvfqdpvgsqtf 56  
QY 1 MEKFMAEFGQGVQTPFLSESNRYKISAGSCPLSTAGSPSVKVFQDPVGSQTF 56  
  
RESULT 2  
ID W76368 standard; Protein; 1693 AA.

```

AC W7638;
AD 03-DEC-1998 (first entry)
DE Hepatitis E virus hollow particle protein #1.
KW Hollow particle protein; virus; antibody; detection; immunoassay;
OS Hepatitis virus. Location/Qualifiers
FH Key
FT Protein 1..1693
FT /note= "Partial sequence"
FT JI0234383-A.
PD 08-SEP-1998.
PF 28-FEB-1997; 062445.
PR 28-FEB-1997; JP-062445.
PA (KOKU-) KOKURITSU YOB0 EISEI KENKYUSHO.
DR WPI: 98-535037/46.
DR N-PSDB: V61687.
PT Hepatitis E virus hollow particle poly-peptide(s) and nucleic acids
PT encoding it - useful for more accurate acid hybridisation
PT using immuno-assays and nucleic acid hybridisation
PS Claim 10; Page 17-24; 29pp; Japanese.
CC This sequence represents a Hepatitis E viral hollow particle protein.
CC This polypeptides can be used to raise antibodies to detect HEV
CC infection in samples, e.g. by immuno-assay based techniques, and the
CC nucleic acid can be used for the same in nucleic acid hybridisation
CC assays. The polypeptides and nucleic acids allow more accurate
CC detection of HEV than previously possible.
SQ Sequence 1693 AA;

Query Match 14.1%; Score 96; DB 35; Length 1693;
Best Local Similarity 26.9%; Pred. No. 4.7le+00;
Matches 18; Conservative 21; Mismatches 25; Indels 3; Gaps 3;

Db 572 frtsfdvgavleangperynlsfdaogstnaagpfspftvaasaaglevryvvaagldhrav 631
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 8 FGOGYVQTPLFSESNSRYRKISAGSCPLSTAGPSYVKFDNPVGSGT-F-SAGLHLR-V 64

Db 632 fapgvsp 638
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 65 FDPSTGA 71

RESULT 3
ID P94617 standard; protein; 520 AA.
AC P94617;
DT 21-JUN-1990 (first entry)
DE Neutral protease encoded by npr gene.
KE Protease; expression systems; subtilin; neutral protease; ds.
OS Bacillus amyloliquefaciens.
FH Key Location/Qualifiers
FT protein 221..520
FT /note="Mature"
FT region 27..520
FT /note="PRO-"
FT US4801537-A.
PD 31-JAN-1989.
PF 29-MAR-1985; 717800.
PR 29-MAR-1985; US-717800.
PA (GENE-) Genex Corp.
PI Nagarajan V, Rhodes CS, Banner CDB;
DR WPI: 89-053639/07.
DR N-PSDB: N91114.
PT Vectors for expression of polypeptide(s) in Bacillus -
PT contg. promoter and regulatory regions which control expression
PT and secretion of protease(s) in Bacillus.
PS Disclosure; p: English.
CC Claimed replicon comprises a promoter and regulatory regions, capable of
CC expressing alkaline and neutral protease genes.
SQ Sequence 520 AA;

Query Match 13.5%; Score 92; DB 1; Length 520;
Best Local Similarity 27.8%; Pred. No. 9.4le+00;
Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps 2;

```





```

RESULT 8
ID W75773 standard; Protein: 623 AA.
AC W75773.
DT 02-DEC-1998 (first entry)
DE Amino acid sequence of lepidoteran-active HD573 toxin.
KW HD573 toxin; PCR; primer; amplification; Bacillus thuringiensis; probe;
KW lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;
KW Helicoverpa zea; hybridisation.
OS Bacillus thuringiensis.
FN W09840490-A1.
PN 17-SEP-1998.
PF 13-MAR-1998; 005081.
PR 13-MAR-1997; US-040512.
PA (MYCO ) MYCOGEN CORP.
PI Muller-Cohn J, Narva KE, Schnepf HE;
DR WPI: 98-506734/43.
DR N-PSDB: V52610.
PT New insecticidal Bacillus thuringiensis toxins - useful for
PT controlling lepidopteran pests, especially Ostrinia nubilalis,
PT Heliothis virescens and Helicoverpa zea
PS Claim 14; Pages 28-30; 50pp; English.
CC This is the amino acid sequence of a novel Bacillus thuringiensis toxin
CC used in the method of the invention, to control lepidopteran pests.
CC The new toxins are useful as pesticides, especially for the control of
CC Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The
CC polynucleotide coding sequences are useful for recombinant expression
CC of the toxins and the primers, together with probes derived from the
CC new sequences, are useful for the identification and characterisation
CC of novel genes that encode pesticidal toxins.
CC Sequence 623 AA;

Query Match 12.7%; Score 87; DB 35; Length 623;
Best Local Similarity 18.2%; Pred. No. 2.21e+01;
Matches 12; Conservative 26; Mismatches 26; Indels 2; Gaps 2;

Db 502 fisekyngdglrfelsttarytirngnsgnynlyrvssigsstirvtngrytanv 561
QY 12 YTPPELSESNRYRKIS-IGSCPLSTAGPSYVKF-QDNPVGSQTFSGAGLHLRVPDPT 69
Db 562 ntntnn 567
QY 70 GALVDS 75

RESULT 9
ID R06621 standard; protein: 239 AA.
AC R06621.
DT 09-JAN-1991 (first entry)
DE Hybrid (1.3-1.4) pre-beta-glucanase.
KW Hybrid pre-beta-glucanase; glucans; beer; animal feed; poultry.
OS Bacillus amyloliquefaciens, Bacillus macerans.
FN Key Location/Qualifiers
FH domain 1..129
FT /label=amino terminal of beta-amyloliquefaciens
FT domain 133..236
FT /label=carboxyl-terminal of B.macerans
PN W09009436-A.
PD 23-AUG-1990.
PF 16-FEB-1990; DK0044.
PR 16-FEB-1989; DD-325800.
PR 04-AUG-1989; DK-003848.
PA (CARL-) CARLSBERG A/S.
PA (DEAK ) AKAD WISSENSCHAFT DDR.
PI Borris R, Hofemeister J, Thomsen KK, Olsen O, Vonwettstein D;
DR WPI: 90-275129/36.
DR N-PSDB: Q05832.
PT New thermostable (1.3-1.4)-beta-glucanase - prepd. using hybrid
PT gene obt'd. using Bacillus amyloliquefaciens and B.macerans genes
PS Disclosure: page 26; 84pp; English.
CC This hybrid protein is encoded by the beta-glucanase-H1 gene.
CC Following processing of the signal peptide the mature protein
CC is produced, comprising the amino terminus of the amylolique-
CC faciens beta-glucanase and the carboxyl-terminal half of the

```

---

```

CC B.macerans beta-glucanase. This hybrid protein is thermostable
CC and hydrolyses beta-glycosidic linkages in (1,3-1,4)-beta-glucans.
CC Reducing sugars are obt'd. at high temps. and thus this enzyme can
CC be used in the mir. of food prods., esp. beer and animal feed (eg
CC for feeding poultry). See also Q05833.
SQ Sequence 239 AA;

Query Match 12.6%; Score 86; DB 1; Length 239;
Best Local Similarity 25.5%; Pred. No. 2.62e+01;
Matches 14; Conservative 20; Mismatches 18; Indels 3; Gaps 3;

Db 74 laltpsynkfcgcnrsvtggyglyevrm-kpakntgivssfftytgtegt 127
QY 36 LSTAGPSYVKFQ-DNPVGSQTFSGAGLH-LRVEDPSTGALVDSKSYAFSTNDTTS 88

RESULT 10
ID W81519 standard; Protein: 1693 AA.
AC W81519.
DT 02-FEB-1999 (first entry)
DE Hepatitis E virus (HEV) ORF-1 protein.
KW Hepatitis E virus; HEV; SAR-55; diagnostic agent; vaccine; antibody;
KW passive immunisation; open reading frame; ORF.
OS Hepatitis E virus.
FN Key Location/Qualifiers
FH Misc_difference 1238 /note= "encoded by CAG"
FT Misc_difference 1244 /note= "encoded by CAG"
FT Misc_difference 1352 /note= "encoded by GTG"
FT Misc_difference 1662 /note= "encoded by GGC"
PN W09846761-A1.
PD 22-OCT-1998.
PF 09-APR-1998; U07418.
PR 11-APR-1997; US-840316.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Emerson SU, Purcell RH, Robinson RA, Tsarev SA;
DR WPI: 98-586733/48.
DR N-PSDB: V71604.
PT New hepatitis E virus DNA from Pakistani strain SAR-55 - used for,
PT e.g. developing products for diagnosis of, and vaccination against
PT hepatitis E virus infection
PS Disclosure: Pages 11-15; 204pp; English.
CC This represents a hepatitis E virus (HEV) open reading frame (ORF)-1
CC protein encoded by a DNA sequence designated SAR-55. SAR-55 also encodes
CC HEV ORF-2 and ORF-3 proteins. A host organism transformed or transfected
CC with a recombinant expression vector containing the SAR-55 nucleic acid
CC can be used to produce the HEV proteins, especially ORF-2 protein. The
CC recombinant HEV proteins can be used as diagnostic agents and as vaccines
CC for use against HEV infection. The detection of antibodies specific for
CC HEV can be used for the diagnosis of infection and diseases caused by
CC HEV, and for monitoring the progression of such disease. Such methods are
CC also useful for monitoring the efficacy of therapeutic agents during the
CC course of treatment of HEV infection and disease in a mammal. The
CC antibodies can be used for detection or for passive immunisation of
CC mammals.
CC Sequence 1693 AA;

Query Match 12.6%; Score 86; DB 37; Length 1693;
Best Local Similarity 25.4%; Pred. No. 2.62e+01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 frtsfvdgavletngperhnlfsdsgtmaagpfslytaasaaglevryvaagldhrav 631
QY 8 FQGGYVQTFPLSESNRYRKIS-IGSCPLSTAGPSYVKFQDNPVGSQTF-SAGLHLR-V 64
Db 632 fapgvsp 638
QY 65 FDPSTGA 71

```



QY 1 1: :  
65 FDPSTGA 71

## RESULT 14

ID R91813 standard; Protein; 1693 AA.  
AC R91813;  
DE 26-NOV-1996 (first entry)  
DT Hepatitis E virus strain SAR-55 ORF-1.  
KW Hepatitis E virus; HEV; SAR-55 strain; enteric transmission;  
KW structural region; antigen; detection; antibody; vaccine;  
KW immunisation; infection.  
OS Hepatitis E virus.  
FH Key  
FT Location/Qualifiers  
FT misc\_difference 1238  
FT misc\_difference 1455..1693  
FT /note= "10 bp nucleic acid sequence TGGTNTTYGA  
has to be inserted between nucleotides  
4390..4391 of T27394 before these amino  
acid residues can be decoded"  
PN WO9610580-A2.  
PD 11-APR-1996.  
PF 03-OCT-1995; U13102.  
PR 03-OCT-1994; US-316765.  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PI Emerson SU, Purcell RH, Tsarev SA;  
DR WPI: 96-209320/21.  
DR N-PSDB; T27394.

Isolated and purified hepatitis E virus strain SAR-55 DNA - encodes  
antigenic protein useful in diagnosis, prophylaxis and treatment of  
hepatitis E virus infection  
Disclosure; Pages 9-13; 121pp; English.  
CC The present sequence is the protein prod. of ORF-1 from the  
CC hepatitis E virus (HEV) strain SAR-55, which was implicated in an  
CC enterically transmitted non-A, non-B hepatitis in Pakistan. The  
CC protein encoded by the structural region of the virus (i.e. ORF-2),  
CC which is capable of forming HEV like particles, is useful for the  
CC detection of HEV antibodies (pref. IgG or IgM) in blood, plasma,  
CC sera, cerebrospinal fluid, tissue, urine or pleural fluid. The  
CC protein, and anti-HEV antibodies generated using the protein, can  
CC also be used in vaccines for immunising an animal against HEV  
CC infection. The protein is identified as a band of greater than  
CC 50 kD following SDS-PAGE of cell lysates of insect cells infected  
CC with a HEV ORF-2 contg. baculovirus, i.e. the claimed recombinant  
CC expression vectors ppIC9-1779, -1780 and -1781.  
SQ Sequence 1693 AA;

Query Match 12.6%; Score 86; DB 19; Length 1693;

Best Local Similarity 25.4%; Pred. No. 2.62e+01;

Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 frtsfdgavletngperhnlstdasqstmaagpslityaasaaglevryvaagldhrav 631

QY 8 FGQGYVOTPLSESNVRYKISITAGSCPLSTAGPSYVKFQDNVPVGSQT-F-SAGLHLR-Y 64

Db 632 fapgvsp 638

QY 1 1: :  
65 FDPSTGA 71

## RESULT 15

ID W75774 standard; Protein; 633 AA.  
AC W75774;  
DE 02-DEC-1998 (first entry)  
DT Amino acid sequence of lepidoteran-active HD525 toxin.  
KW HD525 toxin; PCR; primer; amplification; Bacillus thuringiensis; probe;  
KW lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;  
KW Helicoverpa zea; hybridisation.  
OS Bacillus thuringiensis.  
PN WO9840490-A1.  
PD 17-SEP-1998.  
PF 13-MAR-1998; U05081.

PR 13-MAR-1997; US-040512.

PA (MYCO ) MYCOGEN CORP.

PI Muller-Cohn J, Narva KE, Schnepf HE;

DR WPI: 98-506734/43.

DR N-PSDB; V52611.

PT New insecticidal Bacillus thuringiensis toxins - useful for  
controlling lepidopteran pests, especially Ostrinia nubilalis,

PT Heliothis virescens and Helicoverpa zea

PS Claim 11; Pages 32-34; 50pp; English.

CC This is the amino acid sequence of a novel Bacillus thuringiensis toxin  
used in the method of the invention, to control lepidopteran pests.

CC The new toxins are useful as pesticides, especially for the control of

CC Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The

CC polynucleotide coding sequences are useful for recombinant expression

CC of the toxins and the primers, together with probes derived from the

CC new sequences, are useful for the identification and characterisation

CC of novel genes that encode pesticidal toxins.

SQ Sequence 633 AA;

Query Match 12.4%; Score 85; DB 35; Length 633;

Best Local Similarity 23.7%; Pred. No. 3.10e+01;

Matches 14; Conservative 20; Mismatches 23; Indels 2; Gaps 2;

Db 511 fisekfgnqsdslrfegsnttarytlrgrngsnvnylrvssignstlrvtgrvvytas 569

QY 12 YVQTTPFLSESNVRYKIS-ITAGSCPLSTAGPSYVKF-QDNVPVGSQTFSAGLHLRVFDP 68

Search completed: Wed Sep 1 16:20:34 1999

Job time : 36 secs.





```

CLASSIFICATION #superfamily NADH dehydrogenase (ubiquinone) chain 11  
KEYWORDS iron-sulfur protein; membrane-associated complex;  
metallopotein; mitochondrion; NAD; oxidoreductase  
FEATURE 1-23  
#domain transit peptide (mitochondrion) #status  
predicted #label TNP\  
#product NADH dehydrogenase (ubiquinone) 75K chain  
#status experimental #label MAT  
SUMMARY #length 727 #molecular-weight 79442 #checksum 9285  
  
Query Match 12.7%; Score 87; DB 2; Length 727;  
Best Local Similarity 27.3%; Pred. No. 1.66e+00;  
Matches 21; Conservative 18; Mismatches 34; Indels 4; Gaps 4;  
  
Db 183 IRFASEIAGVDDLTGGRGN-DMQVGTYIEKMFSELNGNIDICPVGALT-SKPXAFTA 240  
QY 24 VRKISIAIGSCPLSTAGPSVKFQDNPVGSQTFSAGLHLRVFDP-STGALVDKSIFYAFST 82  
Db 241 RPWETRKTESIDVMDAV 257  
QY 83 SN-DTTSAAFFVFMNSL 98  
  
RESULT 8  
ENTRY A29091 #type complete  
TITLE licheninase (EC 3.2.1.73) beta - Bacillus amyloliquefaciens  
ALTERNATE_NAMES 1,3-1,4-beta-D-glucan 4-gluconohydrolase; beta-gluconase;  
lichenase  
ORGANISM #formal name Bacillus amyloliquefaciens  
DATE 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change  
ACCESSIONS A29091  
REFERENCE A91564  
#authors Hofemeister, J.; Kurtz, A.; Borriess, R.; Knowles, J.  
#journal Gene (1986) 49:177-187  
#title The beta-gluconase gene from Bacillus amyloliquefaciens shows  
extensive homology with that of Bacillus subtilis.  
#cross-references MUID:87192007  
#accession A29091  
#molecule_type DNA  
#residues 1-239 #label HOF  
#cross-references GB:M15674; NID:g143009; PID:g143010  
#experimental_source strain BZ07/78  
GENETICS  
#gene bglA  
CLASSIFICATION #superfamily licheninase  
KEYWORDS glycosidase; hydrolase; polysaccharide degradation  
SUMMARY #length 239 #molecular-weight 26928 #checksum 1611  
  
Query Match 12.6%; Score 86; DB 2; Length 239;  
Best Local Similarity 25.5%; Pred. No. 2.24e+00;  
Matches 14; Conservative 20; Mismatches 18; Indels 3; Gaps 3;  
  
Db 74 LALTSPYNKFDCGSRNVOTYGYGLYEVRM-KPAKNWTGVSVFFYTGTETP 127  
QY 36 LSTAGPSYXVFQ-DNPVGSQTFSAGLH-LRVFDPSTGALVDKSIFYAFSTNDTTS 88  
  
RESULT 9  
ENTRY I39956 #type complete  
TITLE neutral proteinase (EC 3.4.24.-) - Bacillus amyloliquefaciens  
ORGANISM #formal name Bacillus amyloliquefaciens  
DATE 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change  
ACCESSIONS I39956  
REFERENCE I39956  
#authors Shimada, H.; Honjo, M.; Mita, I.; Nakayama, A.; Akaoka, A.;  
Manabe, K.; Furutani, Y.  
#journal J. Biotechnol. (1985) 2:75-85  
#title The nucleotide sequence and some properties of the neutral  
protease gene of Bacillus amyloliquefaciens.  
#accession I39956  
#status preliminary; translated from GS/EMBL/DDB
```





```

Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scofione, F.; Sekiguchi, J.;
Sekowska, A.; Serror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitenegger, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession B69594
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-242 ##label KUN
#cross-references GB:Z99124; GB:AL009126; NID:g2636442; PID:e1184632;
PID:g2636453
#experimental_source strain 168
REFERENCE
A93526
Murphy, N.; McConnell, D.J.; Cantwell, B.A.
Nucleic Acids Res. (1984) 12:5355-5367
The DNA sequence of the gene and genetic control sites for
the excreted B. subtilis enzyme beta-glucanase.
#cross-references MUID:8427222
#accession A22914
#molecule_type DNA
#residues 1-203, 'L', 205-242 ##label MUR
#cross-references EMBL:X00754; NID:g39818; PID:g685236
#experimental_source strain C120
#note Lys and CCA for residue 82 as Leu
REFERENCE
A90026
Tezuka, H.; Yuuki, T.; Yabuuchi, S.
Agric. Biol. Chem. (1989) 53:2335-2339
Construction of a beta-glucanase hyperproducing Bacillus
subtilis using the cloned beta-glucanase gene and a
multi-copy plasmid.
#accession A90026
#molecule_type DNA
#residues 1-23, 'S', 25-82, 'S', 84-242 ##label TEZ
#cross-references DDBJ:D00518; NID:g216243; PID:g216244
#experimental_source strain Y-25, clone pLE100
REFERENCE
A90027
Yuuki, T.; Tezuka, H.; Yabuuchi, S.
Agric. Biol. Chem. (1989) 53:2341-2346
Purification and some properties of two enzymes from a
beta-glucanase hyperproducing strain, Bacillus subtilis
HL-25
#contents
#note Source was hyperproducing strain HL-25 with gene from strain
Y-25
the amino ends of the mature forms of E-1 and E-2 are
pyroglutamic acid and glutamine, respectively
bglS
catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in
lichenin and other 1,3- and 1,4-beta-linked polysaccharides
#superfamily licheninase
#classification extracellular protein; glycosidase; hydrolase; polysaccharide
degradation; pyroglutamic acid
#feature
1-28 , #domain signal sequence #status predicted #label SIG\
29-242 #product licheninase #status predicted #label MAT\
29 #modified site pyrrolidone carboxylic acid (Gln) (in
mature form) (partial) #status experimental
#length 242 #molecular-weight 27268 #checksum 4439
SUMMARY

```

```

Query Match 12.4%; Score 85; DB 1; Length 242;
Best Local Similarity 25.5%; Pred. No. 3.03e+00;
Matches 14; Conservative 20; Mismatches 18; Indels 3; Gaps 3;

Db 77 LATSPAYNKPDCGERSVQTYGYLYEVRM-KPAKNTGIVSFFTYTGTGTP 130
|: : : : : | | | | | | | | | | | | | | | | | | | | | |
QY 36 LSTAGPSYVKFO-DNPVGSQTFSAGLH-LRVFDPSTGALVDKSKYAFSTSDNTTS 88

RESULT 13
ENTRY A61046 #type complete
TITLE ecdysone-induced membrane protein IMP-E3 - fruit fly
(Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
16-Feb-1997
ACCESSIONS A61046
REFERENCE A61046
#authors Moore, J.T.; Fristrom, D.; Hammonds, A.S.; Fristrom, J.W.
#journal Dev. Genet. (1990) 11:299-309
#title Characterization of IMP-E3, a gene active during imaginal
disc morphogenesis in Drosophila melanogaster.
#accession A61046
#status preliminary
#molecule_type mRNA
#residues 1-331 ##label MOO
GENETICS
#gene FlyBase:ImpE3
#cross-references FlyBase:FBgn0001255
KEYWORDS membrane protein
SUMMARY #length 331 #molecular-weight 36583 #checksum 8221

Query Match 12.4%; Score 85; DB 2; Length 331;
Best Local Similarity 31.9%; Pred. No. 3.03e+00;
Matches 15; Conservative 12; Mismatches 18; Indels 2; Gaps 2;

Db 190 LDNFLRLYDDNGRAAFGSESAMDRWSTASIAGKRRVPPTKP-YVDF 235
|: : : : : | | | | | | | | | | | | | | | | | | | | | |
QY 1 MEKFAEFGQGVQVPFLSESNVRYKI-SIAGSCPLSTAGPSYVKF 46

RESULT 14
ENTRY C56617 #type complete
TITLE cfaC protein precursor - Escherichia coli plasmid NTP113
ORGANISM #formal_name Escherichia coli
DATE 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change
20-Mar-1998
ACCESSIONS C56617
REFERENCE A56617
#authors Jordt, B.J.; Willshaw, G.A.; van der Zeijst, B.A.; Gaastra,
W.
#journal DNA Seq. (1992) 2:257-263
#title The complete nucleotide sequence of region 1 of the CFA/I
fimbrial operon of human enterotoxigenic Escherichia coli.
#cross-references MUID:9232981
#accession C56617
#status preliminary
#molecule_type DNA
#residues 1-869 ##label JOR
#cross-references GB:M55661; NID:g145507; PID:g145510
#experimental_source enterotoxigenic strain, CFA/I-ST plasmid NTP113
#note sequence extracted from NCBI backbone (NCBIN:108960,
NCBIP:108971)
GENETICS
#gene cfaC
#genome plasmid
SUMMARY #length 869 #molecular-weight 97830 #checksum 9755

Query Match 12.4%; Score 85; DB 2; Length 869;
Best Local Similarity 32.4%; Pred. No. 3.03e+00;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;

Db 158 AFIOSQTINLSDSGYKRLSISGNSALGITDTSY 191

```



(TM)

Result	Query			ID	Description	Pred. No.
	No.	Score	Match			
1	300	43.9	65	Y15A_BP74	HYPOTHETICAL 7.3 KD PR	5.56e-433
2	102	14.9	672	NOO3_PARDE	NADH-UBIQUINONE OXIDOR	3.05e-093
3	100	14.6	1693	POLN_HEVYI	NON-STRUCTURAL POLYPRO	6.34e-03
4	92	13.5	521	NPRE_BACAM	BACILLOLYSIN PRECURSOR	1.08e-01
5	89	13.0	438	MAEL_SCHPO	MALIC ACID TRANSPORT P	3.01e-01
6	87	12.7	703	YAIL_RHLSN	HYPOTHETICAL 76.2 KD P	5.88e-01
7	87	12.7	727	NUAM_HUMAN	NADH-UBIQUINONE OXIDOR	5.88e-01
8	87	12.7	727	NUAM_BOVIN	NADH-UBIQUINONE OXIDOR	5.88e-01
9	86	12.6	239	GUB_BACAM	BETA-GLUCANASE PRECURS	8.18e-01
10	86	12.6	457	ENB8_PICGL	LATE EMBRYOGENESIS ABU	8.18e-01
11	86	12.6	932	YNB4_CAEEL	HYPOTHETICAL 104.4 KD	8.18e-01
12	86	12.6	1693	POLN_HEVYA	NON-STRUCTURAL POLYPRO	8.18e-01
13	86	12.6	1693	POLN_HEVPA	NON-STRUCTURAL POLYPRO	8.18e-01
14	85	12.4	242	GUB_BACSU	BETA-GLUCANASE PRECURS	1.14e+00
15	85	12.4	869	CFAC_ECOLI	CFA/I FIMBRIAL SUBUNIT	1.14e+00
16	84	12.3	243	GUB_BACLI	BETA-GLUCANASE PRECURS	1.14e+00
17	83	12.2	171	YAL4_PHUV	HYPOTHETICAL AL4 PROTE	1.57e+00
18	83	12.2	578	SHC_MOUSE	SHC TRANSFORMING PROTE	2.17e+00
19	83	12.2	583	SHC_HUMAN	SHC TRANSFORMING PROTE	2.17e+00
20	83	12.2	611	YX45_MYCTU	HYPOTHETICAL 67.2 KD P	2.17e+00
21	83	12.2	960	KKIT_CHICK	MAST/STEM CELL GROWTH	2.17e+00
22	83	12.2	1325	YAE6_SCHPO	HYPOTHETICAL 145.8 KD	2.17e+00
23	82	12.0	182	Y391_HAEIN	HYPOTHETICAL PROTEIN H	2.99e+00

```

RN  SEQUENCE OF 657-672 FROM N.A.
RX  MEDLINE: 93136200.
RT  "DNA sequencing of the seven remaining structural genes of the gene
RT  cluster encoding the energy-transducing NADH-quinone oxidoreductase
RT  of Paracoccus denitrificans."
RL  BIOCHEMISTRY 32:968-981(1993)
CC  -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC  -!- COFACTOR: MAY BIND TWO 4FE-4S CLUSTER AND ONE 2FE-2S CLUSTER.
CC  -!- SUBUNIT: COMPOSED OF 14 DIFFERENT SUBUNITS.
CC  -!- SIMILARITY: BELONGS TO THE COMPLEX I 75 KD SUBUNIT FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: M84572; G150601; -.
DR  PIR: S23948; S23948.
DR  PROSITE: PS00641; COMPLEX1_75K_1; 1.
DR  PROSITE: PS00642; COMPLEX1_75K_2; 1.
DR  PROSITE: PS00643; COMPLEX1_75K_3; 1.
DR  PIR: PF00111; fer2; 1.
DR  PIR: PF00384; molybdopterin; 1.
KW  OXIDOREDUCTASE; NAD; UBIQUINONE; IRON-SULFUR; 4FE-4S.
FT  INIT_MET 0
FT  METAL 25 25 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 36 36 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 47 47 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 50 50 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 109 109 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 112 112 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 118 118 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 157 157 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT  METAL 160 160 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT  METAL 163 163 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT  METAL 207 207 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ  SEQUENCE 672 AA; 73028 MW; 40DB761A CRC32;

Query Match 14.9%; Score 102; DB 1; Length 672;
Best Local Similarity 26.0%; Pred. No. 3.05e-03;
Matches 20; Conservative 24; Mismatches 29; Indels 4; Gaps 4;

Db 164 VRFTEVAGITQMGQTGRG-EDSEITSYLNQTLNESMGNIIDLCIPGVALY-SRPIYFTA 221
QY 24 VYKISAGSCPLSTAGPSYKFDQNPVGSQTFSAAGLHLRVFDP-STGALYDSKSYAFST 82
Db 222 RPWELTKTESIDVMDAL 238
QY 83 SN-DTTSAAFSFMNSL 98

RESULT 3
ID POLNHEVY STANDARD; PRT; 1693 AA.
AC Q04610;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE NON-STRUCTURAL PROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE
DE (EC 2.7.7.48); HELICASE].
OS HEPATITIS E VIRUS (STRAIN MYANMAR) (HEV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;
CC CALICIVIRUS.
CC [1]
RN SEQUENCE FROM N.A.
RP MEDLINE: 93227573.
RX AYE T.T., UCHIDA T., MA M.Z., IIDA F., SHIKATA T., ICHIKAWA M.,
RX RIKIHISA T., WINN K.;
RT "Sequence and gene structure of the hepatitis E virus isolated from
```

```

RT MYANMAR."
RL VIRUS GENES 7:95-109(1993).
CC -!- HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF ENTERICALLY
CC TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL: D10330; G221705; -.
KW  POLYPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; HELICASE;
KW  ATP-BINDING.
FT  NP_BIND 975 982 ATP (POTENTIAL).
SQ  SEQUENCE 1693 AA; 185215 MW; FFCB786D CRC32;

Query Match 14.6%; Score 100; DB 1; Length 1693;
Best Local Similarity 26.9%; Pred. No. 6.34e-03;
Matches 18; Conservative 22; Mismatches 24; Indels 3; Gaps 3;

Db 572 FRTSFVGVAVLEANGPERYNLSFDASOSTMAAGPFSLTYAASAGLEVRYVAAGLDHRAV 631
QY 8 FGQGVVQTFPLSESNVRYKISAGSCPLSTAGPSYKFDQNPVGSQI-F-SAGLHLR-V 64
Db 632 FAPGVSP 638
QY 65 FDPSTGA 71

RESULT 4
ID NPBE_BACAM STANDARD; PRT; 521 AA.
AC P06832;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE BACILLOLISIN PRECURSOR (EC 3.4.24.28) (NEUTRAL PROTEASE).
GN NPR.
OS BACILLUS AMYLOLIQUEFACIENS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
CC BACILLUS.
CC [1]
RN SEQUENCE FROM N.A.
RP STRAIN-ATCC 23844;
RX MEDLINE: 85006739.
RX VASANTHA N., THOMPSON L.D., RHODES C., BANNER C., NAGLE J.,
RX FILIPULA D.;
RT "Genes for alkaline protease and neutral protease from Bacillus
RT amyloliquefaciens contain a large open reading frame between the
RT regions coding for signal sequence and mature protein."
RL J. BACTERIOL. 159:811-819(1984).
CC -!- FUNCTION: THERMOLABILE EXTRACELLULAR ZINC METALLOPROTEASE.
CC -!- CATALYTIC ACTIVITY: SIMILAR, BUT NOT IDENTICAL, TO THAT OF
CC THERMOLYSIN.
CC -!- COFACTOR: BINDS AND REQUIRES A ZINC ATOM, WHICH IS ESSENTIAL FOR
CC PROTEOLYTIC ACTIVITY. BINDS FOUR CALCIUM IONS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M4 (ZINC METALLOPROTEASE);
CC ALSO KNOWN AS THE THERMOLYSIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL: K02497; G143249; -.
DR  PIR: A25415; HYBSN.
DR  PROSITE: PS00142; ZINC_PROTEASE; 1.
```

[illegible]

Pfam: PF00099; zn-protease; 1.  
DR HSP: P00800; ITRL.  
KW HYDROLASE; METALLOPROTEASE; ZINC; CALCIUM; ZINCOGEN; SIGNAL.

SIGNAL  
FT PROPEP 1 27 POTENTIAL.  
FT PROPEP 28 221 ACTIVATION PEPTIDE.  
FT CHAIN 222 521 BACILLOLYSIN.  
FT METAL 364 364 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 365 365 BY SIMILARITY.  
FT METAL 368 368 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 449 449 PROTON DONOR (BY SIMILARITY).  
SQ SEQUENCE 521 AA; 5684 MW; BC0147D4 CRC32;

Query Match 13.5%; Score 92; DB 1; Length 521;  
Best Local Similarity 27.8%; Pred. No. 1.08e-01;  
Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps 2;

Db 240 ISSESGKYLRLSKPTQIITVDLONREYNLP-GTIVSTTNQFSSQRRA 292  
YQ 36 LTAGPSIV-KPDNPVGSQTFSAGHLRVDFDPSTGALVDSKYAFSTSDNTTS 88

RESULT 5  
ID MAEL\_SCHPO STANDARD; PRN: 438 AA.  
AC P50537:  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE MALIC ACID TRANSPORT PROTEIN (MALATE PERMEASE).  
GN MAEL.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMICOTA; ARCHIASCOMYCETES;  
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;  
CC SCHIZOSACCHAROMYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 96353430.  
RT GROBLER J., BAUER F., SUBDEN R.E., VAN VUUREN H.J.J.;  
RA "The mael gene of Schizosaccharomyces pombe encodes a permease for malate and other C4 dicarboxylic acids.";  
R1 YEAH 11:1485-1491(1995).  
RL -! FUNCTION: PERMEASE FOR MALATE AND OTHER C4 DICARBOXYLIC ACIDS.  
CC -! SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -! SIMILARITY: TO M.JANNA-SCHII MJ0576 AND MJ0762.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

EMBL: U21002; GI171119;  
DR TRANSFORM: TRANSMEMBRANE.  
KW TRANSFORM 37 57 POTENTIAL.  
FT TRANSFORM 37 57 POTENTIAL.  
FT TRANSFORM 65 85 POTENTIAL.  
FT TRANSFORM 106 126 POTENTIAL.  
FT TRANSFORM 140 160 POTENTIAL.  
FT TRANSFORM 172 192 POTENTIAL.  
FT TRANSFORM 205 225 POTENTIAL.  
FT TRANSFORM 242 262 POTENTIAL.  
FT TRANSFORM 288 308 POTENTIAL.  
FT TRANSFORM 321 341 POTENTIAL.  
FT TRANSFORM 353 373 POTENTIAL.  
SQ SEQUENCE 438 AA; 49304 MW; 011F7FB6 CRC32;

Query Match 13.0%; Score 89; DB 1; Length 438;  
Best Local Similarity 39.1%; Pred. No. 3.01e-01;  
Matches 18; Conservative 11; Mismatches 12; Indels 5; Gaps 5;

Db 248 VCPAFAIS-GLAI-I-NIARGAM-CSRPVFVGANSSELVLGVSTFM 289  
YQ 11: :||| ||| | : : : ||| : ||| : : ||| ||| :



```

SQ SEQUENCE 457 AA; 51019 MW; AE7CB4CD CRC32;

Query Match 12.6%; Score 86; DB 1; Length 457;
Best Local Similarity 34.0%; Pred. No. 8.18e-01;
Matches 17; Conservative 11; Mismatches 20; Indels 2; Gaps 2;

Db 232 LGANILVRYLGEVAGNCPLSGAVSLCHPFF-NLVIADDFHKGLGFNNVYD 280
| : ||| :|||:|||| : | : : : ||| : | : |
QY 18 LSENSVRYKISITAGSCPLSTAGPSYVKFQDNPVGSQTSAGLHLR-VFD 66

RESULT 11
ID YMB4-CAEEL STANDARD; PRT; 932 AA.
AC Q03601;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 104.4 KD PROTEIN F54G8.4 IN CHROMOSOME III.
GN F54G8.4.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SCERENITEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE OF 1-574 FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN K.,
RA SULSTON J., THIERRY-MTEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans";
RL NATURE 368:32-38(1994).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z19155; G6715; -
DR PIR; S28275; S28275.
DR WORMPEP; F54G8.4; CE05942.
DR PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
DR PFAM; PF00097; zf-C3HC4; 1.
DR HSP; P9590; IBOF.
DR KW HYPOTHETICAL PROTEIN; ZINC-FINGER; DNA-BINDING; NUCLEAR PROTEIN.
FT ZN_FING 43 83
FT C3HC4-TYPE.
SQ SEQUENCE 932 AA; 104438 MW; 45C877B4 CRC32;

Query Match 12.6%; Score 86; DB 1; Length 932;
Best Local Similarity 38.2%; Pred. No. 8.18e-01;
Matches 13; Conservative 9; Mismatches 10; Indels 2; Gaps 2;

Db 282 TETEGLEYLRNFPADAELENRK-LRFSAGDDAA 314
| : ||| ||| :||| : | : : ||| : | : |
QY 55 TFSAGLH-LRVFDPSTGALVDKSYAFSTSNDDT 87

RESULT 12
ID POLN.HEYBU STANDARD; PRT; 1693 AA.
IC P2032A.

```





```

RP SEQUENCE OF 29-63.
RA YUUKI T., TEZUKA H., YABUCHI S.;
RT "Purification and some properties of two enzymes from a beta-glucanase
RL hyperproducing strain, Bacillus subtilis HL-25.";
RN AGRIC. BIOL. CHEM. 53:2341-2346(1989).
RP SEQUENCE FROM N.A.
RX MEDLINE: 93329981.
RA JORDI B.J.A.M., WILLSHAW G.A., VAN DER ZEIJST B.A.M., GAASTRA W.;
RT "The complete nucleotide sequence of region 1 of the CFA/I fimbrial
RL DNA SEQ. 2:257-263(1992).
CC "The complete nucleotide sequence of region 1 of the CFA/I fimbrial
CC DNA SEQ. 2:257-263(1992).
CC -!- FUNCTION: MAY SERVE AS ANCHOR FOR THE FIMBRIAE IN THE OUTER
CC MEMBRANE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X00754; G685236;
DR EMBL: D00518; D1000859;
DR EMBL: D83026; D1012364;
DR EMBL: Z46862; G590674;
DR EMBL: Z28340; G580948;
DR EMBL: Z99124; E1184632;
DR PIR: A22914; LXBS.
DR PIR: JU0110; JU0110.
DR SUBTILIST: BG10476; BGLS.
DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
DR PFAM: PF00722; glycosyl_hydro9; 1.
DR HSP: P27051; IGBG.
KW HYDROLASE; GLYCOSIDASE; SIGNAL.
FT SIGNAL 1 28
FT CHAIN 29 242
FT MOD_RES 29 29
FT ACT_SITE 133 133
FT ACT_SITE 137 137
FT DISULFID 60 89
FT VARIANT 24 24
FT VARIANT 83 83
FT VARIANT 204 204
FT SEQUENCE 242 AA; 27268 MW; 3B62B79A CRC32;
Query Match 12.4%; Score 85; DB 1; Length 242;
Best Local Similarity 25.5%; Pred. No. 1.14e+00;
Matches 14; Conservative 20; Mismatches 18; Indels 3; Gaps 3;
Db 77 LALTPAYNKFCCGNSRQVYGYLYEVRM-KPAKNTGIVSFFTYGTGDTGP 130
QY 36 LSTAGPYVYKFO-DNPVGSQTFSGALH-LRVDPSTGALVDKSKYAFSTNDTTS 88

RESULT 15
ID CFAC-ECOLI STANDARD; PRT; 869 AA.
AC P25733;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DE 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE CFA/I FIMBRIAL SUBUNIT C PRECURSOR (COLONISATION FACTOR ANTIGEN I
DE SUBUNIT C).
GN CFAC.
OS ESCHERICHIA COLI.
OG PLASMID NTP513.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ENTEROTOXIGENIC;
RX MEDLINE: 89330163.
RA HAMERS A.M., PEL H.J., WILLSHAW G.A., KUSTERS J.G.,
RA VAN DER ZEIJST B.A.M., GAASTRA W.;

```

```

RT "The nucleotide sequence of the first two genes of the CFA/I fimbrial
RT operon of human enterotoxigenic Escherichia coli.";
RN MICROB. PATHOG. 6:297-309(1989).
RP SEQUENCE FROM N.A.
RX MEDLINE: 93329981.
RA JORDI B.J.A.M., WILLSHAW G.A., VAN DER ZEIJST B.A.M., GAASTRA W.;
RT "The complete nucleotide sequence of region 1 of the CFA/I fimbrial
RL DNA SEQ. 2:257-263(1992).
CC "The complete nucleotide sequence of region 1 of the CFA/I fimbrial
CC DNA SEQ. 2:257-263(1992).
CC -!- FUNCTION: MAY SERVE AS ANCHOR FOR THE FIMBRIAE IN THE OUTER
CC MEMBRANE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M55661; G145510;
KW ANTIGEN; SIGNAL; FIMBRIA; OUTER MEMBRANE; PLASMID.
FT SIGNAL 1 22
FT CHAIN 23 869
FT SEQUENCE 869 AA; 97830 MW; 7AF76347 CRC32;
Query Match 12.4%; Score 85; DB 1; Length 869;
Best Local Similarity 32.4%; Pred. No. 1.14e+00;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;
Db 158 AFTQSTINLSDSGYKRLSIGNSALGITDTSY 191
QY 11 GYVQPFLESNSVRVK-ISIAGSCPLSTAGPSY 43

Search completed: Wed Sep 1 16:18:21 1999
Job time : 13 secs.

```

Result No.	Query			ID	Description	Pred. No.
	Score	Match	Length			
1	101	14.8	1577	2	Q54178	1.85e-02
2	100	14.6	1693	14	Q81876	2.53e-02
3	96	14.1	1114	14	Q41088	9.70e-02
4	92	13.5	229	10	P93830	3.53e-01
5	92	13.5	229	14	Q65574	3.53e-01
6	92	13.5	300	14	Q65581	3.53e-01
7	92	13.5	337	2	Q44647	3.53e-01
8	91	13.3	229	10	Q49162	3.53e-01
9	90	13.2	3443	14	Q11979	4.85e-01
10	89	13.0	566	2	Q84354	6.15e-01
11	89	13.0	921	5	P90770	9.10e-01
12	88	12.9	180	10	Q49984	9.10e-01
13	88	12.9	1546	2	Q45996	9.10e-01
14	87	12.7	239	2	Q45691	1.24e+00
15	86	12.6	521	2	Q44671	1.24e+00
16	86	12.6	574	5	Q18470	1.69e+00
17	86	12.6	974	5	Q23510	2.30e+00
18	86	12.6	1693	14	Q59410	2.30e+00
19	86	12.6	1693	14	Q81344	2.30e+00
20	86	12.6	1693	14	Q89444	2.30e+00



RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-JURA;  
RA SCHWYZER M.;  
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; 248053; G971327; -;  
KW EMBL; AJ004801; E1187337; -;  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 299 AA; 32379 MW; A3897D4A CRC32;

Query Match 13.5%; Score 92; DB 14; Length 299;  
Best Local Similarity 47.2%; Pred. No. 3.53e-01;  
Matches 17; Conservative 9; Mismatches 6; Indels 4; Gaps 4;

Db 45 PRFVCEVREIPAGPTFTSSSITHLRV-EPSTGALL 79

QY 41 PSYV-KFQDNVGSQTF-SAGL-HLRVDFPSTGALV 73

RESULT 6  
ID Q65581 PRELIMINARY; PRT; 300 AA.

AC Q65581;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE UL7 POLYPEPTIDE.  
GN UL7

OS BOVINE HERPESVIRUS TYPE 1.  
OC VIRUSES: DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;  
OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-SCHOENBOEREN;

RX MEDLINE; 96135223

RA SCHMITT J., KEIL G.M.;

RT "Identification and characterization of the bovine herpesvirus 1 UL7 gene and gene product which are not essential for virus replication in cell culture."

RL J. VIROL. 70:1091-1099(1996).

DR EMBL; X91751; G1006630; -;

SQ SEQUENCE 300 AA; 32450 MW; 86377347 CRC32;

Query Match 13.5%; Score 92; DB 14; Length 300;  
Best Local Similarity 47.2%; Pred. No. 3.53e-01;  
Matches 17; Conservative 9; Mismatches 6; Indels 4; Gaps 4;

Db 46 PRFVCEVREIPAGPTFTSSSITHLRV-EPSTGALL 80

QY 41 PSYV-KFQDNVGSQTF-SAGL-HLRVDFPSTGALV 73

RESULT 7  
ID Q44647 PRELIMINARY; PRT; 337 AA.

AC Q44647;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
DE NEUTRAL PROTEASE (FRAGMENT).

OS BACILLUS AMYLOLIQUEFACIENS.

OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;

OC BACILLUS.

RN [1]

RP SEQUENCE FROM N.A.

RA YOSHIMURA K., MIYAZAKI T., NAKAHAMA K., KIKUCHI M.;

RL TAKEDA KENYUSHO HO 44:42-50(1985).

DR EMBL; M64815; G142890; -;

KW PROTEASE.

FT NON\_TER 337 337

SQ SEQUENCE 337 AA; 36910 MW; 5A9BC52B CRC32;

Query Match 13.5%; Score 92; DB 2; Length 337;  
Best Local Similarity 27.8%; Pred. No. 3.53e-01;

Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps 2;

Db 240 ISSESGKVVLRDLSPKTCQIITYDLQREYNLP-CTLVSSSTNQFTTSSQRAA 292

QY 36 LSTAGPSYV-KFQDNVGSQTF-SAGLHLRVDFPSTGALVDSKSYAFSTSNQITS 88

RESULT 8

ID O49162 PRELIMINARY; PRT; 229 AA.

AC O49162;

DT 01-JUN-1998 (TREMBLREL. 06, CREATED)

DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE IAA17/AXR3-1 PROTEIN.

GN AXR3.

OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).

OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;

OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;

OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;

RX MEDLINE; 98146427.

RA ROUSE D., MACKAY P., STIRNBERG P., ESTELLE M., LEYSER O.;

RT "Changes in auxin response from mutations in an AUX/IAA gene."

RL SCIENCE 279:1371-1373(1998).

DR EMBL; AF040632; G2921758; -;

SQ SEQUENCE 229 AA; 25304 MW; 2D88B037 CRC32;

Query Match 13.3%; Score 91; DB 10; Length 229;  
Best Local Similarity 22.1%; Pred. No. 4.85e-01;

Matches 17; Conservative 22; Mismatches 33; Indels 5; Gaps 5;

Db 89 VRSYRKNWVSCQKSGGPEAAAFVKVMDGAPYLKIDLRMYK-SYDELSNALSNNFSS 147

QY 24 VR-YKISAGSCLPSTAGPSYVYKQDNVGSQTF-SAGLHLRVDFPSTGALVDS-KS-YA- 79

Db 148 FTMGRKHGEGEEMIDFWN 164

QY 80 FSTSDNTTSAAFVSPFN 96

RESULT 9

ID O11979 PRELIMINARY; PRT; 3443 AA.

AC O11979;

DT 01-JUL-1997 (TREMBLREL. 04, CREATED)

DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE POLYPROTEIN.

OS MAIZE CHLOROTIC DWARF WAIKAVIRUS.

OC VIRUSES: SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; SEQUIVIRIDAE;

OC WAIKAVIRUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-TENNESSEE (TN);

RX MEDLINE; 97296968.

RA REDDICK B.B., HABERA L.F., LAW M.D.;

RT "Nucleotide sequence and taxonomy of maize chlorotic dwarf virus within the family Sequiviridae."

RL J. GEN. VIROL. 78:1165-1174(1997).

DR EMBL; U67839; G2160664; -;

DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.

DR PFAM; PF00910; RNA\_helicase; 1.

KW POLYPROTEIN.

SQ SEQUENCE 3443 AA; 387659 MW; 7E60C323 CRC32;

Query Match 13.2%; Score 90; DB 14; Length 3443;  
Best Local Similarity 31.8%; Pred. No. 6.65e-01;

Matches 14; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

Db 1193 VDSQERRYGYLRIWHPDNGSLDEGVFAMSTNLEPPFGAFVKY 1236

QY 51 VGSQTF-SAGLHLRVDFPSTGALVDSKSYAFSTSDNTTSAAFVSF 94





\*\*\*\*\*  
 M P S R E L E H  
 \*\*\*\*\*  
 (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Wed Sep 1 16:23:00 1999; MasPar time 4.06 Seconds  
 Tabular output not generated.  
 52.355 Million cell updates/sec

Title: >PCT-US99-13024-2  
 Description: (8-17) from PCTUS9913024.pap (8 of 12)  
 Perfect Score: 82  
 Sequence: 1 FGQGYVOTPF 10  
 Scoring table: PAM 150  
 Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-geneseq35  
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21 22:part22 23:part23  
 24:part24 25:part25 26:part26 27:part27 28:part28  
 29:part29 30:part30 31:part31 32:part32 33:part33  
 34:part34 35:part35 36:part36 37:part37 38:part38  
 39:part39

Statistics: Mean 17.449; Variance 54.450; scale 0.320

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	53	64.6	30	8 R40072	Hib OMP p2 peptide OM	7.02e+01
2	53	64.6	43	8 R40095	Hib OMP p1-p2 hybrid	7.02e+01
3	53	64.6	341	1 R07043	P3 gene product of Ha	7.02e+01
4	53	64.6	341	1 R07042	P3 gene product of Ha	7.02e+01
5	53	64.6	342	13 R66880	H. influenzae protein	7.02e+01
6	53	64.6	342	13 R66879	H. influenzae protein	7.02e+01
7	53	64.6	361	13 R66877	H. influenzae p2 enco	7.02e+01
8	53	64.6	361	1 P90599	P2 gene product of Ha	7.02e+01
9	53	64.6	361	1 P90098	P2 antigen of Haemoph	7.02e+01
10	53	64.6	363	13 R66878	pNV-2 fusion construc	7.02e+01
11	53	64.6	390	23 W21678	Haemophilus influenza	7.02e+01
12	51	62.2	796	17 R97616	Cephalosporin amidohy	1.13e+02
13	51	62.2	5059	30 W52846	A. mediterranei rifam	1.13e+02
14	50	61.0	613	32 W55964	Serratia marcescens S	1.43e+02
15	50	61.0	613	7 R36774	Serratia marcescens	1.43e+02
16	49	59.8	211	38 W85101	Thyroid hormone recep	1.81e+02

17	49	59.8	212	10	R53543	Thyroid hormone recep	1.81e+02
18	49	59.8	621	29	W55645	H. pylori ORF 06ep106	1.81e+02
19	49	59.8	637	30	W55739	H. pylori ORF 06ep106	1.81e+02
20	49	59.8	645	37	W73030	Helicobacter pylori 7	1.81e+02
21	49	59.8	652	37	W73029	Helicobacter pylori 7	1.81e+02
22	49	59.8	782	37	W79701	B. subtilis ECB deacy	1.81e+02
23	49	59.8	782	37	W79704	B. subtilis ECB deacy	1.81e+02
24	49	59.8	782	37	W79702	B. subtilis ECB deacy	1.81e+02
25	49	59.8	782	37	W79703	B. subtilis ECB deacy	1.81e+02
26	49	59.8	782	37	W79705	B. subtilis ECB deacy	1.81e+02
27	49	59.8	888	22	W16494	DNA polymerase which	1.81e+02
28	48	58.5	659	15	R89265	Heparinase-III.	2.28e+02
29	48	58.5	1480	3	R13302	CFTR L1077P.	2.28e+02
30	47	57.3	17	26	W39017	Peptide resembling an	2.88e+02
31	47	57.3	122	12	R67992	Aprotinin.	2.88e+02
32	47	57.3	127	12	R67993	Aprotinin.	2.88e+02
33	47	57.3	166	38	W73233	Porcine interferon-ga	2.88e+02
34	47	57.3	166	38	W83904	Porcine interferon-ga	2.88e+02
35	47	57.3	786	4	R26001	Aculeacin A acylase.	2.88e+02
36	47	57.3	876	25	W34675	Human mammary carcino	2.88e+02
37	47	57.3	882	25	W34673	Human mammary carcino	2.88e+02
38	47	57.3	914	12	R71100	Protein-tyrosine-kin	2.88e+02
39	47	57.3	919	13	R75502	Human mammary carcino	2.88e+02
40	47	57.3	919	13	R75504	Human mammary carcino	2.88e+02
41	47	57.3	919	25	W34672	Human mammary carcino	2.88e+02
42	46	56.1	17	28	W38104	Peptide recognition u	3.62e+02
43	46	56.1	319	39	W67803	Human retinol dehydro	3.62e+02
44	46	56.1	649	9	R47504	Pyrococcus furiosus a	4.54e+02
45	45	54.9	218	39	W83371	Streptococcus pneumon	4.54e+02

## ALIGNMENTS

RESULT 1  
 ID R40072 standard; peptide; 30 AA.  
 AC R40072;  
 DT 04-FEB-1994 (first entry)  
 DE Hib OMP p2 peptide OMP2-2 (53-81).  
 KW Haemophilus influenzae; type b; Hib; outer membrane protein; P1; P2;  
 KW P6; vaccine; antibody; detection; lipoglycopeptide conjugate;  
 KW immunogen.  
 OS Synthetic.  
 FH Key  
 FT misc\_difference 30  
 FT Location/Qualifiers  
 FT /note= "May be absent"  
 PN W09315205-A.  
 PD 05-AUG-1993.  
 PF 03-FEB-1993; CA0041.  
 PR 03-FEB-1992; GB-002219.  
 PA (CONN-) CONNAUGHT LAB LTD.  
 PI Chong P, Kandil A, Klein MH, Sia C;  
 DR WPI; 93-258681/32.  
 PT Synthetic Haemophilus influenzae conjugate vaccine - comprising  
 PT T-helper cell determinants and B-cell epitope(s) linked to  
 PT synthetic oligo:saccharide(s)  
 PS Table 2; Page 49; 99pp; English.  
 CC The sequences given in R40053-101 are peptide fragments derived from  
 CC the Haemophilus influenzae type b (Hib) outer membrane proteins P1,  
 CC P2 and P6. These peptides may be used in a vaccine against Hib  
 CC infection and antibodies against these peptides may be used in test  
 CC kits to detect H. influenzae in a sample. The vaccine may further  
 CC comprise an immunogenic or immunostimulatory molecule or the peptides  
 CC may be modified with lipids, or linked to synthetic PRP as synthetic  
 CC lipoglycopeptide conjugates to produce alternative vaccines. This  
 CC sequence is represented as it is given in the specification.  
 SQ Sequence 30 AA;

Query Match 64.6%; Score 53; DB 8; Length 30;  
 Best Local Similarity 50.0%; Pred. No. 7.02e+01;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 Db 6 yaqgyletrf 15  
 ::|||::|

Qy 8 FGQGYVQTPF 17

RESULT 2  
 ID R40095 standard; peptide; 43 AA.  
 AC R40095; 1994 (first entry)  
 DT 04-FEB-1994 (first entry)  
 DE Hib OMP P1-P2 hybrid peptide IP13-2P2.  
 KW Haemophilus influenzae; type b; Hib; outer membrane protein; P1; P2;  
 KW P6; vaccine; antibody; detection; lipoglycopeptide conjugate;  
 KW immunogen.  
 OS Synthetic.  
 FT Key Location/Qualifiers  
 FT region 1..14  
 FT /label= C-P1  
 FT region 15..43  
 FT /label= CHIBP2  
 PN W09315205-A.  
 PD 05-AUG-1993. CA0041.  
 PF 03-FEB-1993; GB-002219.  
 PR 03-FEB-1992; GB-002219.  
 PA (CONN-) CONNAUGHT LAB LTD.  
 PI Chong P, Kandil A, Klein MH, Sia C;  
 PI WPI: 93-258681/32.  
 DR Synthetic Haemophilus influenzae conjugate vaccine - comprising  
 PT T-helper cell determinants and B-cell epitope(s) linked to  
 PT synthetic oligo:saccharide(s)  
 PS Table 11; Page 59; 99pp; English.  
 CC The sequences given in R40053-101 are peptide fragments derived from  
 CC the Haemophilus influenzae type b (Hib) outer membrane proteins P1,  
 CC P2 and P6. These peptides may be used in a vaccine against Hib  
 CC infection and antibodies against these peptides may be used in test  
 CC kits to detect H. influenzae in a sample. The vaccine may further  
 CC comprise an immunogenic or immunostimulatory molecule or the peptides  
 CC may be modified with lipids, or linked to synthetic PRP as synthetic  
 CC lipoglycopeptide conjugates to produce alternative vaccines.  
 CC Sequence 43 AA.  
 SQ

Query Match 64.6%; Score 53; DB 8; Length 43;  
 Best Local Similarity 50.0%; Pred. No. 7.02e+01;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 20 yaqgyletrf 29  
 Qy 8 FGQGYVQTPF 17

RESULT 3  
 ID R07043 standard; protein; 341 AA.  
 AC R07043; 1990 (first entry)  
 DT 04-DEC-1990 (first entry)  
 DE P3 gene product of Haemophilus strain 8358 (OMP subtype 6U).  
 KW P3 gene; strain 1H; influenza; vaccine; ds.  
 OS Haemophilus influenzae.  
 PN EP-378929-A.  
 PD 25-JUL-1990.  
 PF 22-DEC-1989; 313573.  
 PR 23-DEC-1988; GB-030124.  
 PR 01-FEB-1989; GB-002178.  
 PA (CONN-) CONNAUGHT LAB LTD.  
 PI Munson RS, Tolan RW, Chong P, Fahim R, McVerry P, Klein M;  
 PI WPI: 90-225607/30.  
 DR N-PSDB; Q06120.  
 PT Gene coding for protein P2 of Haemophilus influenzae type-B -  
 PT used for developing vaccines for protection against disease  
 PT caused by the organism.  
 PS Claim 1; Page 8; 15pp; English.  
 CC Peptide may be used as a vaccine to the disease caused by H1  
 CC type b, as carrier for conjugation to oligosaccharide derived  
 CC from Haemophilus. Protein may be used with other haptens as  
 CC T-cell dependant antigen and carrier.  
 CC Sequence 341 AA;  
 SQ

Query Match 64.6%; Score 53; DB 1; Length 341;  
 Best Local Similarity 50.0%; Pred. No. 7.02e+01;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 58 yaqgyletrf 67  
 Qy 8 FGQGYVQTPF 17

RESULT 4  
 ID R07042 standard; protein; 341 AA.  
 AC R07042;  
 DT 04-DEC-1990 (first entry)  
 DE P3 gene product of Haemophilus strain Durot (OMP subtype 2L).  
 KW P3 gene; strain 1H; influenza; vaccine; ds.  
 OS Haemophilus influenzae.  
 PN EP-378929-A.  
 PD 25-JUL-1990.  
 PF 22-DEC-1989; 313573.  
 PR 23-DEC-1988; GB-030124.  
 PR 01-FEB-1989; GB-002178.  
 PA (CONN-) CONNAUGHT LAB LTD.  
 PI Munson RS, Tolan RW, Chong P, Fahim R, McVerry P, Klein M;  
 PI WPI: 90-225607/30.  
 DR N-PSDB; Q06119.  
 PT Gene coding for protein P2 of Haemophilus influenzae type-B -  
 PT used for developing vaccines for protection against disease  
 PT caused by the organism.  
 PS Claim 1; Page 8; 15pp; English.  
 CC Peptide may be used as a vaccine to the disease caused by H1  
 CC type b, as carrier for conjugation to oligosaccharide derived  
 CC from Haemophilus. Protein may be used with other haptens as  
 CC T-cell dependant antigen and carrier.  
 CC Sequence 341 AA;  
 SQ

Query Match 64.6%; Score 53; DB 1; Length 341;  
 Best Local Similarity 50.0%; Pred. No. 7.02e+01;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 58 yaqgyletrf 67  
 Qy 8 FGQGYVQTPF 17

RESULT 5  
 ID R66880 standard; Protein; 342 AA.  
 AC R66880;  
 DT 22-AUG-1995 (first entry)  
 DE H. influenzae protein P2 encoded by pNV-6.  
 KW Haemophilus influenzae type b; outer membrane protein; P2; pNV-6.  
 OS Haemophilus influenzae type b.  
 PN W09503069-A.  
 PD 02-FEB-1995.  
 PF 22-JUL-1994; U08326.  
 PR 23-JUL-1993; US-096181.  
 PA (NABA-) NORTH AMERICAN VACCINE INC.  
 PI Liang S, Pullen JK, Soper TS, Tai JY;  
 PI WPI: 95-075024/10.  
 DR N-PSDB; Q84357.  
 PT Production of Haemophilus influenzae HibP2 protein - for the  
 PT prevention of bacterial meningitis in animals  
 PS Claim 15; Figure 7; 63pp; English.  
 CC To clone P2 into expression vector pET-11a (Novagen pET System  
 CC Manual) a 1.1kb fragment from full length P2 (in pNV-1) (see  
 CC Q84354) was generated using PCR and oligos Q85350 and Q84353.  
 CC Oligo Q85350 allowed the mature P2 protein to be cloned into the NdeI  
 CC site of pET 11a. Q84353 incorporated a BamHI site. The P2 fragment  
 CC thus generated was digested with NdeI and BamHI, purified and  
 CC ligated into pET11a previously cut with NdeI-BamHI. This resulted  
 CC in a mature P2 construct - pNV-6. The rest of the plasmid is  
 CC identical to pET 11a.  
 CC Sequence 342 AA;  
 SQ



```

Query Match      64.6%; Score 53; DB 13; Length 342;
Best Local Similarity 50.0%; Pred. No. 7.02e+01;
Matches      5; Conservative      4; Mismatches      1; Indels      0; Gaps      0;

Db      59 yaqgyetrf 68
      ::::|::|
QY      8 FGQGYVQTFF 17

RESULT
ID      R66877 standard; Protein; 342 AA.
AC      R66877;
DE      22-AUG-1995 (first entry)
KW      H. influenzae protein P2 encoded by pNV-3.
OS      Haemophilus influenzae type b; outer membrane protein; P2; pNV-3.
PN      W09503069-A.
PD      02-FEB-1995.
PF      22-JUL-1994; U08326.
PR      23-JUL-1993; US-096181.
PA      (NAVA-) NORTH AMERICAN VACCINE INC.
PI      Liang S, Pullen JK, Soper TS, Tai JY;
DR      WPI: 95-075024/10.
DR      N-PSDB: Q84356.
PT      Production of Haemophilus influenzae HibP2 protein - for the
PT      prevention of bacterial meningitis in animals
PS      Claim 13; Figure 6; 63pp; English.
CC      pNV-3 contains mature P2 contg. a Met and the N-terminus.
CC      To clone P2 into expression vector pET-17b (Novagen PET System
CC      Manual) the original P2 gene (in pNV-1) (see Q84354) was modified
CC      that allowed the mature protein to be cloned into the NdeI site
CC      of pET 17b, thus producing the mature P2. The oligo designed for
CC      this is Q84350. The extraneous 3' sequences were eliminated by
CC      introducing xhoI site about 40 bp from the translation
CC      stop codon. This oligo (Q84352) contd. an xhoI site to allow it to
CC      be cloned into the xhoI site of pET-17b. PCR was used to generate a
CC      1.1 kb fragment from the full length P2 (pNV-1) with the 5' oligo
CC      contg. the NdeI site and the 3' oligo contg. the xhoI site.
CC      This fragment was digested with NdeI and xhoI, purified and ligated
CC      into the NdeI-xhoI digested pET-17b. This yielded the mature P2
CC      construct pNV-3.
SQ      Sequence 342 AA;

Query Match      64.6%; Score 53; DB 13; Length 342;
Best Local Similarity 50.0%; Pred. No. 7.02e+01;
Matches      5; Conservative      4; Mismatches      1; Indels      0; Gaps      0;

Db      59 yaqgyetrf 68
      ::::|::|
QY      8 FGQGYVQTFF 17

RESULT
ID      R66877 standard; Protein; 361 AA.
AC      R66877;
DE      22-AUG-1995 (first entry)
KW      H. influenzae P2 encoded by pNV-1 cDNA.
OS      Haemophilus influenzae type b; outer membrane protein; P2.
PN      W09503069-A.
PD      02-FEB-1995.
PF      22-JUL-1994; U08326.
PR      23-JUL-1993; US-096181.
PA      (NAVA-) NORTH AMERICAN VACCINE INC.
PI      Liang S, Pullen JK, Soper TS, Tai JY;
DR      WPI: 95-075024/10.
DR      N-PSDB: Q84354.
PT      Production of Haemophilus influenzae HibP2 protein - for the
PT      prevention of bacterial meningitis in animals
PS      Example; Figure 4; 63pp; English.
CC      Total genomic DNA was isolated from H. influenzae type b strain Eagan
CC      and used as a template for two P2 specific oligos in a PCR. The 5',

```

```

CC      oligo (Q84348) was designed to be 40 bp 5' of the ATG start codon.
CC      The 3' oligo (Q84349) was designed to be 300 bp 3' of the stop codon.
CC      Both oligos contd. SalI restriction sites. The amplified DNA was
CC      used to transform E.coli and the resulting colonies isolated and
CC      analysed. Clones contg. a 1.4 kb fragment were chosen for DNA
CC      sequence analysis. One clone, designated pNV-1, was found to be
CC      identical to the published sequence for Hib strain Minn A (Munson,
CC      R. and Tolan, R.W, Infection & Immunity 57:88-94, Jan. 1989). The
CC      synthetic oligos used to sequence the DNA are indicated in the FT.
CC      The direction of sequencing was 5'-3'. The rest of pNV-1 is
CC      identical to pUC18. The lac promotor is adjacent to the lower
CC      SalI site.
SQ      Sequence 361 AA;

Query Match      64.6%; Score 53; DB 13; Length 361;
Best Local Similarity 50.0%; Pred. No. 7.02e+01;
Matches      5; Conservative      4; Mismatches      1; Indels      0; Gaps      0;

Db      78 yaqgyetrf 87
      ::::|::|
QY      8 FGQGYVQTFF 17

RESULT
ID      R05999 standard; protein; 361 AA.
AC      R05999;
DE      04-DEC-1990 (first entry)
KW      P2 gene product of Haemophilus strain 1H.
OS      P2 gene; strain 1H; influenza; vaccine; ds.
OS      Haemophilus influenzae.
PN      EP-378929-A.
PD      25-JUL-1990.
PF      22-DEC-1989; 313573.
PR      23-DEC-1988; GB-030124.
PR      01-FEB-1989; GB-002178.
PA      (CONN-) CONNAUGHT LAB LTD.
PI      Munson RS, Tolan RW, Chong P, Fahim R, McVerry P, Klein M;
DR      WPI: 90-225607/30.
DR      N-PSDB: Q05372.
PT      Gene coding for protein P2 of Haemophilus influenzae type-B -
PT      used for developing vaccines for protection against disease
PT      caused by the organism.
PS      Claim 1; Page 8; 15pp; English.
CC      Peptide may be used as a vaccine to the disease caused by H1
CC      type b, as carrier for conjugation to oligosaccharide derived
CC      from Haemophilus. Protein may be used with other haptens as
CC      T-cell dependant antigen and carrier.
SQ      Sequence 361 AA;

Query Match      64.6%; Score 53; DB 1; Length 361;
Best Local Similarity 50.0%; Pred. No. 7.02e+01;
Matches      5; Conservative      4; Mismatches      1; Indels      0; Gaps      0;

Db      78 yaqgyetrf 87
      ::::|::|
QY      8 FGQGYVQTFF 17

RESULT
ID      P90098 standard; protein; 362 AA.
AC      P90098;
DE      1-NOV-1989 (first entry)
DE      P2 antigen of Haemophilus influenzae
KW      Haemophilus influenzae P2 antigen; vaccines; pathogenic
KW      type b strains.
OS      Haemophilus influenzae type b
PN      EP-320289-A.
PD      14-JUN-1989.
PF      09-DEC-1988; 311691.
PR      10-DEC-1987; US-249482.
PA      (TEXA) Univ of Texas Syst.
PI      Hansen EJ;
DR      WPI: 89-174562/24.

```

PT DNA encoding Haemophilus influenzae P2 antigen  
 PT - used for producing immunogenic compsns. as vaccines  
 PT and in diagnostic hybridisation assays.

PS Disclosure; fig 6; 20pp; English.

CC P2 antigen of Haemophilus influenzae (see corresp. N90033).

CC Isolated from type b strains, the P2 antigens are used as vaccines

CC and to detect P2-specific RNA or DNA. Segments used as antigens

CC include amino acids 21-361, 175-197, 260-275, 296-311, 33-353,

CC or their functional equivalents.

SQ Sequence 362 AA;

Query Match 64.6%; Score 53; DB 1; Length 362;

Best Local Similarity 50.0%; Pred. No. 7.02e+01;

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 79 yaqgyletrf 88

QY 8 FGQGYVQTFP 17

RESULT 10

ID R66878 standard; Protein; 363 AA.

AC R66878;

DT 22-AUG-1995 (first entry)

DE PNV-2 fusion construct of H. influenzae P2 with T7 gene 10 protein.

KW Haemophilus influenzae type b; outer membrane protein; P2; PNV-2;

KW fusion protein; fusion P-2; gene 10; major capsid protein; T7.

OS Synthetic.

PN W09503069-A.

PD 02-FEB-1995.

PF 22-JUL-1994; U08326.

PR 23-JUL-1993; US-096181.

PA (NANA-) NORTH AMERICAN VACCINE INC.

PI Liang S., Pullen JK, Soper TS, Tai JY;

DR WPI: 95-075024/10.

DR N-PSDB; Q84355.

PT Production of Haemophilus influenzae HibP2 protein - for the

PT prevention of bacterial meningitis in animals

PS Claim 14; Figure 5; 63pp; English.

CC Fusion P2 contains mature outer membrane protein P2 from

CC Haemophilus influenzae type b (hib-P2) with 22 AAs of gene 10 of

CC expression vector (Novagen pET System Manual). To P2 into

CC pET-17b the original P2 gene (in pNV-1) (see Q84354) was modified

CC using PCR. To construct the fusion-P2, an oligo was constructed

CC allowed the mature porin to be cloned into the BamHI site of

CC pET-17b, thus yielding a fusion protein P2 to gene 10. The

CC sequence of this oligo was Q84351. The extraneous 3' sequences were

CC eliminated by introducing xhoI site about 40 bp from the translation

CC stop codon. This oligo (Q84352) contd. an xhoI site to allow it to

CC be cloned into the xhoI site of pET-17b. PCR was used to generate a

CC 1.1 kb fragment from the full length P2 (PNV-1) with the 5' oligo

CC contg. the BamHI site and the 3' oligo contg. the xhoI site. This

CC fragment was digested with BamHI and xhoI, purified and ligated

CC into the BamHI-xhoI digested pET-17b. This yielded the fusion-P2

CC construct PNV-2.

SQ Sequence 363 AA;

Query Match 64.6%; Score 53; DB 13; Length 363;

Best Local Similarity 50.0%; Pred. No. 7.02e+01;

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 80 yaqgyletrf 89

QY 8 FGQGYVQTFP 17

RESULT 11

ID W21678 standard; Protein; 390 AA.

AC W21678;

DT 30-SEP-1997 (first entry)

DE Haemophilus influenzae outer membrane protein P2.

KW Fusion protein; protein delivery; outer membrane protein P2;

KW vaccine; immunogen; respiratory disease.

OS Haemophilus influenzae strain 1479.

FH Key Location/Qualifiers

FT region 22..43

FT /label= Loop-1

FT region 69..85

FT /label= Loop-2

FT region 115..133

FT /label= Loop-3

FT region 155..175

FT /label= Loop-4

FT region 203..248

FT /label= Loop-5

FT region 262..277

FT /label= Loop-7

FT region 303..315

FT /label= Loop-7

FT region 331..364

FT /label= Loop-8

PN W09716207-A1.

PD 09-MAY-1997.

PF 01-NOV-1996; U17698.

PR 31-OCT-1996; US-740644.

PR 02-NOV-1995; US-006168.

PA (UYNV ) UNIV NEW YORK STATE RES FOUND.

PI Kyungcheol Y, Timothy FM;

DR WPI: 97-271883/24.

DR N-PSDB; T72476.

PT New recombinant DNA for producing fusion proteins - comprising

PT microbial pathogen epitope inserted into Haemophilus influenzae P2

PT protein loop 5, used in vaccine

PS Example 1; Page 40-43; 54pp; English.

CC The loop 5 portion of Haemophilus influenzae strain 1479 outer

CC fusion with, and display of, one or more heterologous or homologous

CC peptide sequences. A method is provided for producing a fusion

CC protein comprising a P2 protein in which loop 5, or a portion of

CC it, displays a heterologous or homologous peptide sequence which

CC antigenically mimics an epitope from a microbial pathogen. The

CC fusion protein can be expressed in a host cell (e.g. as a surface-

CC exposed sequence) and used as an immunogen against disease caused

CC by H. influenzae and/or the microbial pathogen. It is esp. used

CC to treat infection by respiratory pathogens.

SQ Sequence 390 AA;

Query Match 64.6%; Score 53; DB 23; Length 390;

Best Local Similarity 50.0%; Pred. No. 7.02e+01;

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 yaqgyletrf 87

QY 8 FGQGYVQTFP 17

RESULT 12

ID R97616 standard; Protein; 796 AA.

AC R97616;

DT 23-AUG-1996 (first entry)

DE Cephalosporin amidohydrolase.

KW Cephalosporin amidohydrolase;

KW 7-aminocephalosporanic acid; gac gene.

OS Pseudomonas vesicularis strain B965.

PN W09616174-A2.

PD 30-MAY-1996.

PF 15-NOV-1995; E04487.

PR 17-NOV-1994; GB-023212.

PA (GLAX ) GLAXO GROUP LTD.

PI Burr KW, Harrison LA, Illing GT, Maishman NJ, Ramsden M;

PI Slade A, Spence D W;

DR WPI: 96-268610/27.

DR N-PSDB; T29215.

PT One-step conversion of cephalosporin C into 7-aminocephalosporanic

PT acid - allows environmentally friendly production of

PT 7-amino:cephalosporanic acid for prod. of cephalosporin-like  
PT antibiotics  
PS Disclosure: Page 32-36; 40pp; English.  
CC A cephalosporin amidohydrolase (R97616) was identified as the  
CC product of the *gac* gene (T29215) of *Pseudomonas vesicularis* strain  
CC B965. The enzyme is capable of converting cephalosporin C and  
CC its derivatives to 7-aminocephalosporanic acid and its  
CC derivatives, building blocks of semisynthetic cephalosporin  
CC antibiotics. The method is cheap, involves simple technology and  
CC is environmentally safe. The enzyme is obtd. by incorporation of  
CC the *gac* gene into a vector and expression in e.g. *Escherichia coli*.  
CC The purified enzyme is pref. immobilised.  
SQ Sequence 796 AA;

Query Match 62.2%; Score 51; DB 17; Length 796;  
Best Local Similarity 100.0%; Pred. No. 1.13e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 83 fgqgyv 88  
| | | | |  
QY 8 FGQGYV 13

RESULT 13  
ID W52846 standard; Protein; 5069 AA.  
AC W52846;  
DT 24-JUL-1998 (first entry)  
DE A. mediterranei rifamycin synthesis gene cluster fragment protein B.  
KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;  
KW polyketide synthase; actinomycete; ansamycin.  
OS Amycolatopsis mediterranei.  
PN WO9807868-A1.  
PD 26-FEB-1998.  
PR 18-AUG-1997; E04495.  
PR 20-AUG-1996; EP-810551.  
PA (NOVS) NOVARTIS AG.  
PI Engel N, Schupp T, Toupet C;  
DR N-PSDB: V21187.  
PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used  
PT to produce rifamycin and rifamycin analogues  
PS Claim 6; Page 126-151; 205pp; English.  
CC The present sequence represents a Amycolatopsis mediterranei rifamycin  
CC synthesis gene cluster ORF B protein from the present invention. The  
CC DNA fragment comprises a DNA region involved directly or indirectly  
CC in the gene cluster responsible for rifamycin synthesis, including  
CC the adjacent DNA regions to the right and left which, by reason of  
CC their function in connection with rifamycin biosynthesis, qualify  
CC as constituents of this rifamycin gene cluster, and functional  
CC fragments, derivatives or constituents of these. The Amycolatopsis  
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used  
CC for producing rifamycin, rifamycin analogues or precursors. It can also  
CC be used for inactivating or modifying genes involved in ansamycin or  
CC rifamycin biosynthesis. The DNA can be used for constructing mutant  
CC actinomycetes strains from which the natural rifamycin or ansamycin  
CC biosynthesis gene cluster has been partly or completely deleted. The  
CC DNA fragment can be used for assembling a library of polyketide  
CC synthases, which can be used for assembling a library of polyketides.  
CC A hybridisation probe of the invention can be used for identifying DNA  
CC fragments involved in the biosynthesis of ansamycins.  
SQ Sequence 5069 AA;

Query Match 62.2%; Score 51; DB 30; Length 5069;  
Best Local Similarity 100.0%; Pred. No. 1.13e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3455 fgqgyv 3460  
| | | | |  
QY 8 FGQGYV 13

RESULT 14  
ID W55964 standard; Protein; 613 AA.

AC W55964;  
DT 21-AUG-1998 (first entry)  
DE Serratia marcescens Sr41 esterase sequence.  
KW Serratia marcescens mutant M-1; FERM BP-4068; esterase;  
KW S marcescens Sr41; FERM BP-467; higher productivity.  
OS Serratia marcescens.  
PN EP-837135-A2.  
PD 22-APR-1998.  
PF 25-NOV-1992; 120078.  
PR 13-APR-1992; JP-137502.  
PR 25-NOV-1991; JP-355440.  
PA (TANA) TANABE SEIYAKU CO.  
PI Shibatani T;  
DR WPI: 98-219117/20.  
DR N-PSDB: V26063; V26064.  
PT Serratia marcescens mutant M-1 - with high esterase productivity  
PS Example 3; Pages 10-13; 15pp; English.  
CC The present sequence represents an esterase of Serratia marcescens.  
CC A S. marcescens mutant was produced and designated M-1 (FERM  
CC BP-4068). The M-1 mutant is obtained by treating S. marcescens  
CC Sr41 (FERM BP-467) with N-methyl-N'-nitro-N-nitrosoguanidine. The M-1  
CC mutant has a two fold higher productivity of esterase than the wild  
CC type organism. The mutant is used for producing M-1 esterase.  
SQ Sequence 613 AA;

Query Match 61.0%; Score 50; DB 32; Length 613;  
Best Local Similarity 60.0%; Pred. No. 1.43e+02;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 36 fdgyhqtgf 45  
| | | | |  
QY 8 FGQGYVQTPF 17

RESULT 15  
ID R36774 standard; Protein; 613 AA.  
AC R36774;  
DT 28-SEP-1993 (first entry)  
DE Serratia esterase.  
KW Sr41; M1; mutation; allele; variation; hydrolysis.  
OS Serratia marcescens.  
PN EP-544250-A.  
PD 02-JUN-1993.  
PF 25-NOV-1992; 120078.  
PR 25-NOV-1991; JP-355440.  
PR 13-APR-1992; JP-137502.  
PA (TANA) TANABE SEIYAKU CO.  
PI Akatsuka H, Komatsubara S, Omori K, Shibatani T;  
DR WPI: 93-177043/22.  
DR N-PSDB: Q42857; Q42858.  
PT New genes encoding esterase from Serratia - are contained in  
PT recombinant vectors and transformed cells with high esterase  
PT productivity, useful as hydrolytic reagent  
PS Disclosure; Page 11-13; 16pp; English.  
CC When S. marcescens Sr41 was subjected to mutagenesis with  
CC N-methyl-N'-nitro-N-nitrosoguanidine and mutants selected on  
CC tributyrin medium, the mutant M1 was identified. M1 contains  
CC an allelic variant of *c* -> *t* at position 9. M1 has 2.5 times  
CC greater esterase prod. than Sr41. The esterase is useful for  
CC carrying out hydrolytic reactions.  
SQ Sequence 613 AA;

Query Match 61.0%; Score 50; DB 7; Length 613;  
Best Local Similarity 60.0%; Pred. No. 1.43e+02;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 36 fdgyhqtgf 45  
| | | | |  
QY 8 FGQGYVQTPF 17

Search completed: Wed Sep 1 16:23:25 1999  
Job time : 25 secs.



\*\*\*\*\*

W P S R E H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Sep 1 16:22:24 1999; MasPar time 5.25 Seconds  
Tabular output not generated. 76.309 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (8-17) from PCTUS9913024.pep (8 of 12)  
Perfect Score: 82  
Sequence: 1 FGQGYVQTPE 10

Scoring table: PAM 150  
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4  
Statistics: Mean 24.385; Variance 32.736; scale 0.745

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query	Match	Length	ID	Description	Pred. No.
1	57	69.5	262	2	S64050	hypothetical protein	1.13e+00	
2	55	67.1	509	2	C69491	probable acid-CoA li	2.77e+00	
3	55	67.1	1036	1	A34755	nitrogen regulatory p	2.77e+00	
4	54	65.9	547	2	A43832	heme-binding protein	4.30e+00	
5	54	65.9	549	2	D64098	major outer membrane	6.66e+00	
6	53	64.6	359	2	D64050	major outer membrane	6.66e+00	
7	53	64.6	361	2	A30542	major outer membrane	6.66e+00	
8	53	64.6	371	2	S68072	major outer membrane	6.66e+00	
9	53	64.6	371	2	S68069	major outer membrane	6.66e+00	
10	53	64.6	392	2	C69572	probable aspartate tr	6.66e+00	
11	53	64.6	936	2	B64567	cytochrome c biogenes	6.66e+00	
12	53	64.6	936	2	H71862	probable cytochrome C	6.66e+00	
13	52	63.4	60	2	JN0741	cellulose 1,4-beta-ce	1.02e+01	
14	52	63.4	484	1	A24994	cellulose 1,4-beta-ce	1.02e+01	
15	52	63.4	1123	2	A44766	defective chorion-1 f	1.02e+01	
16	51	62.2	340	2	G70432	conserved hypothetical	1.57e+01	
17	51	62.2	743	2	S19437	hypothetical protein	1.57e+01	
18	51	62.2	822	2	S26872	DNA-directed RNA poly	1.57e+01	
19	51	62.2	830	1	RNEGB2	DNA-directed RNA poly	1.57e+01	
20	51	62.2	2301	2	T02323	hypothetical protein	1.57e+01	
21	50	61.0	234	2	F64595	membrane fusion prote	2.38e+01	
22	50	61.0	234	2	G71918	probable efflux trans	2.38e+01	
23	50	61.0	489	2	C64984	lysine-specific perme	2.38e+01	

24 50 61.0 612 2 A54282 reversed polarity pro 2.38e+01  
25 50 61.0 613 2 JS0763 triacylglycerol lipas 2.38e+01  
26 50 61.0 915 2 T00728 hypothetical protein 2.38e+01  
27 49 59.8 336 2 PC1139 hypothetical protein 3.60e+01  
28 49 59.8 418 2 F70669 hypothetical protein 3.60e+01  
29 49 59.8 425 2 E71039 hypothetical protein 3.60e+01  
30 49 59.8 608 2 B64610 outer membrane protei 3.60e+01  
31 49 59.8 629 2 E64610 outer membrane protei 3.60e+01  
32 49 59.8 638 2 B71904 probable outer membra 3.60e+01  
33 49 59.8 651 2 E71904 probable outer membra 3.60e+01  
34 49 59.8 659 2 A43734 probable protein kin 3.60e+01  
35 49 59.8 779 2 A57177 NIMA-like protein kin 3.60e+01  
36 49 59.8 786 2 JC1298 aculeacin-A acylase p 3.60e+01  
37 49 59.8 930 2 E64098 DNA-directed DNA poly 3.60e+01  
38 48 58.5 241 2 S77548 lumQ protein - Synec 5.41e+01  
39 48 58.5 402 2 F64389 chromate resistance p 5.41e+01  
40 48 58.5 408 2 C70870 probable lpps protein 5.41e+01  
41 48 58.5 481 2 A26483 hemolytic protein hhd 5.41e+01  
42 48 58.5 532 2 S70844 hemolytic protein hhd 5.41e+01  
43 48 58.5 625 2 H70330 hypothetical protein 5.41e+01  
44 48 58.5 659 2 JC4910 heparitin-sulfate lya 5.41e+01  
45 48 58.5 3473 1 A46112 genome polyprotein - 5.41e+01

## ALIGNMENTS

RESULT 1  
ENTRY S64050 #type complete  
TITLE hypothetical protein YGL046w - yeast (Saccharomycetes cerevisiae)  
ALTERNATE\_NAMES  
ORGANISM #formal\_name Saccharomycetes cerevisiae  
DATE 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 14-Nov-1997  
ACCESSIONS S64050  
REFERENCE S64044  
#authors Feuermann, M.; Potier, S.; Souciet, J.L.  
#submission submitted to the Protein Sequence Database, May 1996  
#accession S64050  
#molecule\_type DNA  
#residues 1-262 ##label FEU  
##cross-references EMBL:Z72568; NID:gl322534; PID:e243943; PID:gl322535; MIPS:YGL046w  
#experimental\_source strain S288C  
GENETICS  
#map\_position 7L  
SUMMARY #length 262 #molecular-weight 29440 #checksum 290  
Query Match 69.5%; Score 57; DB 2; Length 262;  
Best Local Similarity 60.0%; Pred. No. 1.13e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 107 YGQYVKTAF 116  
Qy 8 FGQGYVQTPE 17  
RESULT 2  
ENTRY C69491 #type complete  
TITLE probable acid-CoA ligase (EC 6.2.1.-) fadD8 - Archaeoglobus fulgidus  
ORGANISM #formal\_name Archaeoglobus fulgidus  
DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 12-Feb-1999  
ACCESSIONS C69491  
REFERENCE A69250  
#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Keriavage, A.R.; Graham, D.E.; Kyripides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.;

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiaich, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

#journal Nature (1997) 390:364-370  
 #title The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*.  
 #cross-references MUID:98049343  
 #accession C69491  
 #status preliminary; nucleic acid sequence not shown; translation not shown

##molecule\_type DNA  
 ##residues 1-509 ##label KLE  
 ##cross-references GB:AE000970; GB:AE000782; NID:g2689293; PID:g2648612; TIGR:AF1932

CLASSIFICATION #superfamily acetate--CoA ligase homology  
 KEYWORDS acid-thiol ligase  
 FEATURE 52-499  
 SUMMARY #length 509 #molecular-weight 57035 #checksum 45

Query Match 67.1% Score 55; DB 2; Length 509;  
 Best Local Similarity 60.0%; Pred. No. 2.77e+00;  
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 325 FTGCGYGVVF 334  
 I : I I I I I I  
 QY 8 FGCGYVOTPF 17

RESULT 3  
 ENTRY A34755 #type complete  
 TITLE nitrogen regulatory protein nit-2 - *Neurospora crassa*  
 ORGANISM #formal name *Neurospora crassa*  
 DATE 13-Jul-1990 #sequence\_revision 26-Jul-1996 #text\_change 16-Feb-1997

ACCESSIONS A34755  
 REFERENCE A34755  
 #authors Fu, Y.H.; Marzluf, G.A.  
 #journal Mol. Cell. Biol. (1990) 10:1056-1065  
 #title nit-2, the major nitrogen regulatory gene of *Neurospora crassa*, encodes a protein with a putative zinc finger DNA-binding domain.  
 #cross-references MUID:90158568  
 #accession A34755  
 ##molecule\_type mRNA  
 ##residues 1-1036 ##label FUY  
 ##cross-references GB:M33956

GENETICS 209/2; 335/3  
 #introns  
 CLASSIFICATION #superfamily nitrogen regulatory protein nit-2; GATA-type zinc finger homology  
 KEYWORDS DNA binding; transcription regulation; zinc finger  
 FEATURE 740-793 #domain GATA-type zinc finger homology #label GZF\  
 743-767 #region zinc finger GATA motif  
 SUMMARY #length 1036 #molecular-weight 109295 #checksum 235

Query Match 67.1% Score 55; DB 1; Length 1036;  
 Best Local Similarity 66.7%; Pred. No. 2.77e+00;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 981 FNGYINTP 989  
 I I I I I I  
 QY 8 FGCGYVOTP 16

RESULT 4  
 ENTRY A43832 #type complete  
 TITLE heme-binding protein A precursor - *Haemophilus influenzae* (type b)

ORGANISM #formal name *Haemophilus influenzae*  
 DATE 07-May-1993 #sequence\_revision 07-May-1993 #text\_change 17-Oct-1997  
 ACCESSIONS A43832  
 REFERENCE A43832  
 #authors Hanson, M.S.; Slaughter, C.; Hansen, E.J.  
 #journal Infect. Immun. (1992) 60:2257-2266  
 #title The hbpA gene of *Haemophilus influenzae* type b encodes a heme-binding lipoprotein conserved among heme-dependent *Haemophilus* species.  
 #cross-references MUID:92267636  
 #contents type b  
 #accession A43832  
 ##molecule\_type DNA; protein  
 ##residues 1-547 ##label HAN  
 ##cross-references GB:M88134  
 ##note sequence extracted from NCBI backbone (NCBIN:104242, NCBIP:104243)

GENETICS  
 #gene hbpA  
 CLASSIFICATION #superfamily dipeptide transport protein  
 KEYWORDS heme  
 FEATURE 1-18  
 19-547

SUMMARY #length 547 #molecular-weight 60692 #checksum 7908  
 Query Match 65.9% Score 54; DB 2; Length 547;  
 Best Local Similarity 85.7%; Pred. No. 4.30e+00;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 528 GYQSPF 534  
 I I I I I I  
 QY 11 GYQTPF 17

RESULT 5  
 ENTRY D64098 #type complete  
 TITLE heme-binding protein A precursor - *Haemophilus influenzae* (strain Rd K20)

ORGANISM #formal name *Haemophilus influenzae*  
 DATE 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 24-Oct-1997

ACCESSIONS D64098  
 REFERENCE A64000  
 #authors Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.; FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.  
 #journal Science (1995) 269:496-512  
 #title Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.  
 #cross-references MUID:95350630  
 #accession D64098  
 ##status nucleic acid sequence not shown; translation not shown  
 ##molecule\_type DNA  
 ##residues 1-549 ##label TIGR  
 ##cross-references GB:U32767; GB:LA2023; NID:g1573868; PID:g1573869; TIGR:HI0853

GENETICS  
 #gene hbpA  
 CLASSIFICATION #superfamily dipeptide transport protein  
 KEYWORDS heme  
 FEATURE 1-20  
 21-549

Db 981 FNGYINTP 989  
 I I I I I I  
 QY 8 FGCGYVOTP 16

RESULT 4  
 ENTRY A43832 #type complete  
 TITLE heme-binding protein A precursor - *Haemophilus influenzae* (type b)  
 #domain signal sequence #status predicted #label SIG\  
 #product heme-binding protein A #status predicted #label

```

SUMMARY      #length 549  #molecular-weight 60904  #checksum 3028
Query Match   65.9%; Score 54; DB 2; Length 549;
Best Local Similarity 85.7%; Pred. No. 4.30e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 530 GYVQSPF 536
      ||||:|
QY 11 GYVQTPF 17

RESULT      6
ENTRY      D64050  #type complete
TITLE      major outer membrane protein P2 - Haemophilus influenzae
           (strain Rd KW20)
ORGANISM   #formal_name Haemophilus influenzae
DATE       18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
           24-Oct-1997
ACCESSIONS D64050
REFERENCE   A64000
#authors    Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
           Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.;
           Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
           Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
           Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
           J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
           M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
           D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,
           J.L.; Geoghegan, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
           Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
#journal    Science (1995) 269:496-512
#title      Whole-genome random sequencing and assembly of Haemophilus
           influenzae Rd.
#cross-references MUID:95350630
#accession  D64050
#status     nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues   1-359 #label TIGR
#cross-references GB:U32699; GB:L42023; NID:g1573084; PID:g1573092;
           TIGR:H10139
KEYWORDS    membrane protein
SUMMARY      #length 359  #molecular-weight 39375  #checksum 1781
Query Match   64.6%; Score 53; DB 2; Length 359;
Best Local Similarity 50.0%; Pred. No. 6.66e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 YAOGYLETRF 87
      ::|||:|
QY 8 FQGGYVQTPF 17

RESULT      7
ENTRY      A30542  #type complete
TITLE      major outer membrane protein P2 precursor - Haemophilus
           influenzae (type b)
ORGANISM   #formal_name Haemophilus influenzae
DATE       07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
           20-Mar-1998
ACCESSIONS A30542; A32750; A60116; A60121; S09622
REFERENCE   A30542
#authors    Munson Jr., R.; Tolan Jr., R.W.
#journal    Infect. Immun. (1989) 57:88-94
#title      Molecular cloning, expression, and primary sequence of outer
           membrane protein P2 of Haemophilus influenzae type b.
#cross-references MUID:89079316
#accession  A30542
#molecule_type DNA
#residues   1-361 #label MUN
#cross-references GB:J03359; NID:g148958; PID:g148959
#experimental_source type b
#note       part of this sequence, including the amino end of the

```

```

REFERENCE    A32750  mature protein, was confirmed by protein sequencing
#authors     Hansen, E.J.; Hasemann, C.; Clausell, A.; Capra, J.D.; Orth,
           K.; Moomaw, C.R.; Slaughter, C.A.; Latimer, J.L.; Miller,
           E.E.
#journal     Infect. Immun. (1989) 57:1100-1107
#title       Primary structure of the porin protein of Haemophilus
           influenzae type b determined by nucleotide sequence
           analysis.
#cross-references MUID:89173305
#accession   A32750
#status     not compared with conceptual translation
#molecule_type DNA
#residues    1-361 #label HAN
#experimental_source type b, strain DL42
#note       mature protein, was confirmed by protein sequencing

REFERENCE    A60116
#authors     Hansen, E.J.; Pelzel, S.E.; Orth, K.; Moomaw, C.R.; Radolf,
           J.D.; Slaughter, C.A.
#journal     Infect. Immun. (1989) 57:3270-3275
#title       Structural and antigenic conservation of the P2 porin protein
           among strains of Haemophilus influenzae type b.
#cross-references MUID:90035390
#accession   A60116
#molecule_type protein
#residues    21-42 #label HA2
#experimental_source type b, strains DL42, CH100, DL26, DV102, H234,
           Madigan, OA104, and SL103

REFERENCE    A60121
#authors     Hansen, E.J.; Gonzales, F.R.; Chamberlain, N.R.; Norgard,
           M.V.; Miller, E.E.; Cope, L.D.; Pelzel, S.E.; Gaddy, B.;
           Clausell, A.
#journal     Infect. Immun. (1988) 56:2709-2716
#title       Cloning of the gene encoding the major outer membrane protein
           of Haemophilus influenzae type b.
#cross-references MUID:88330197
#accession   A60121
#molecule_type protein
#residues    21-40 #label HA3
#experimental_source type b, strain DL42

REFERENCE    S09622
#authors     Munson Jr., R.; Bailey, C.; Grass, S.
#journal     Mol. Microbiol. (1989) 3:1797-1803
#title       Diversity of the outer membrane protein P2 gene from major
           clones of Haemophilus influenzae type b.
#cross-references MUID:90158127
#accession   S09622
#status     preliminary; not compared with conceptual translation
#molecule_type DNA
#residues    1-361 #label M02
#keywords     membrane protein
FEATURE      1-20
           21-361
SUMMARY      #domain signal sequence #status predicted #label SIG\
           #product major outer membrane protein P2 #status
           experimental #label MAT
           #length 361 #molecular-weight 39701 #checksum 4550
Query Match   64.6%; Score 53; DB 2; Length 361;
Best Local Similarity 50.0%; Pred. No. 6.66e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 YAOGYLETRF 87
      ::|||:|
QY 8 FQGGYVQTPF 17

RESULT      8
ENTRY      S68072  #type complete
TITLE      major outer membrane protein P2, antigenic variant t1 -
           Haemophilus influenzae
ORGANISM   #formal_name Haemophilus influenzae
DATE       19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change

```

```

24-Oct-1997
S68072: S45437
S68062
Duim, B.
#authors
#submission submitted to the EMBL Data Library, June 1993
#accession S68072
##status preliminary
##molecule_type DNA
##residues 1-371 #label DU1
##cross-references EMBL:X73386; NID:g510590; PID:g510591
##experimental_source isolate t1
REFERENCE
S43699
#authors Duim, B.; van Alphen, L.; Eijk, P.; Jansen, H.M.; Dankert, J.
#journal Mol. Microbiol. (1994) 11:1181-1189
#title Antigenic drift of non-encapsulated Haemophilus influenzae
#title major outer membrane protein p2 in patients with chronic
#title bronchitis is caused by point mutations.
#cross-references MUID:94293786
#accession S45437
##status preliminary
##molecule_type DNA
##residues 174-188;215-234;257-283 #label DUW
##cross-references EMBL:X73386
##experimental_source isolate t1
KEYWORDS
membrane protein
SUMMARY
#length 371 #molecular-weight 40901 #checksum 3994
Query Match 64.6%; Score 53; DB 2; Length 371;
Best Local Similarity 50.0%; Pred. No. 6.66e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 YAOGYLETRF 87
:||||:|
QY 8 FGQGYVQTFF 17

RESULT 9
ENTRY
TITLE
#molecule_type DNA
#residues 1-371 #label DU1
#cross-references EMBL:X73383; NID:g475192; PID:e1192128; PID:g2654292
#experimental_source isolate t2
REFERENCE
S43699
#authors Duim, B.; van Alphen, L.; Eijk, P.; Jansen, H.M.; Dankert, J.
#journal Mol. Microbiol. (1994) 11:1181-1189
#title Antigenic drift of non-encapsulated Haemophilus influenzae
#title major outer membrane protein p2 in patients with chronic
#title bronchitis is caused by point mutations.
#cross-references MUID:94293786
#accession S43706
##status preliminary
##molecule_type DNA
##residues 174-188;215-234;257-283 #label DUW
##cross-references EMBL:X73383
##experimental_source isolate t2
KEYWORDS
membrane protein
SUMMARY
#length 371 #molecular-weight 40945 #checksum 5124
Query Match 64.6%; Score 53; DB 2; Length 371;
Best Local Similarity 50.0%; Pred. No. 6.66e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 78 YAOGYLETRF 87
:||||:|
QY 8 FGQGYVQTFF 17

RESULT 10
ENTRY
TITLE
#molecule_type DNA
#residues 1-392 #label KUN
#cross-references GB:299111; GB:AL009126; NID:g2633699; PID:e1184990;
#experimental_source strain 168
REFERENCE
A33496
#authors Perego, M.; Cole, S.P.; Burbulys, D.; Trach, K.; Hoch, J.A.
#journal J. Bacteriol. (1989) 171:6187-6196
#title Characterization of the gene for a protein kinase which
#title phosphorylates the sporulation-regulatory proteins spo0A
#title and spo0F of Bacillus subtilis.
#cross-references MUID:90036708
#accession C33496
##status
##molecule_type DNA
##residues 1-96, 'S', 99-232, 'T', 234-392 #label PER
##experimental_source strain 168

Db 78 YAOGYLETRF 87
:||||:|
QY 8 FGQGYVQTFF 17

RESULT 10
ENTRY
TITLE
#molecule_type DNA
#residues 1-392 #label KUN
#cross-references GB:299111; GB:AL009126; NID:g2633699; PID:e1184990;
#experimental_source strain 168
REFERENCE
A33496
#authors Perego, M.; Cole, S.P.; Burbulys, D.; Trach, K.; Hoch, J.A.
#journal J. Bacteriol. (1989) 171:6187-6196
#title Characterization of the gene for a protein kinase which
#title phosphorylates the sporulation-regulatory proteins spo0A
#title and spo0F of Bacillus subtilis.
#cross-references MUID:90036708
#accession C33496
##status
##molecule_type DNA
##residues 1-96, 'S', 99-232, 'T', 234-392 #label PER
##experimental_source strain 168

```



```

GENETICS
#gene      pata
CLASSIFICATION
#superfamily aspartate transaminase
#amino transferase
KEYWORDS
SUMMARY
#length 392 #molecular-weight 43448 #checksum 5874

Query Match      64.6%; Score 53; DB 2; Length 392;
Best Local Similarity 50.0%; Pred. No. 6.66e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 351 YGEGYVRLSF 360
      :|::|::|:
QY 8 FGQGYVQTFF 17

RESULT 11
ENTRY
TITLE      B64567 #type complete
            cytochrome c biogenesis protein - Helicobacter pylori (strain
            26695)
ORGANISM   #formal_name Helicobacter pylori
DATE       09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
            10-Oct-1997
ACCESSIONS B64567
REFERENCE   A64520
#authors    Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
            Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
            H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
            J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
            Richardson, D.; Dodson, R.; Khalaf, H.G.; Glodek, A.;
            McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
            Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
            Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
            Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
            W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
            C.M.; Venter, J.C.

#journal    Nature (1997) 388:539-547
#title      The complete genome sequence of the gastric pathogen
            Helicobacter pylori.
#cross-references GB:AE00554; GB:AE00554; NID:92313475; PID:92313481;
            TIGR:HP0378
#accession  B64567
#status     preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues   1-936 #label TOM
#cross-references GB:AE00554; GB:AE00554; NID:92313475; PID:92313481;
            TIGR:HP0378
SUMMARY     #length 936 #molecular-weight 106359 #checksum 8931

Query Match      64.6%; Score 53; DB 2; Length 936;
Best Local Similarity 50.0%; Pred. No. 6.66e+00;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 222 FGSAYIELPF 231
      ||::|::|
QY 8 FGQGYVQTFF 17

RESULT 12
ENTRY
TITLE      H71862 #type complete
            probable cytochrome C-type biogenesis protein - Helicobacter
            pylori (strain J99)
ORGANISM   #formal_name Helicobacter pylori
#variety    strain J99
DATE       12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
            12-Feb-1999
ACCESSIONS H71862
REFERENCE   A71800
#authors    Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
            Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
            B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
            Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
            Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
            G.F.; Trust, T.J.

#journal    Nature (1999) 397:176-180
#title      Genomic sequence comparison of two unrelated isolates of the
            human gastric pathogen Helicobacter pylori.
#cross-references MUID:99120557
#accession  H71862
#status     preliminary
#molecule_type DNA
#residues   1-936 #label ARN
#cross-references GB:AE001528; GB:AE001439; NID:94155575; PID:94155589
#experimental_source strain J99

GENETICS
#gene      jhp1003
SUMMARY     #length 936 #molecular-weight 106452 #checksum 7770

Query Match      64.6%; Score 53; DB 2; Length 936;
Best Local Similarity 50.0%; Pred. No. 6.66e+00;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 222 FGSAYIELPF 231
      ||::|::|
QY 8 FGQGYVQTFF 17

RESULT 13
ENTRY
TITLE      JN0741 #type complete
            hypothetical 6.6K protein - phage SPp1
ALTERNATE_NAMES
            hypothetical protein 50
ORGANISM   #formal_name phage SPp1
DATE       03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change
            17-Mar-1999
ACCESSIONS JN0741; S21439
REFERENCE   JN0729
#authors    Chai, S.; Szepan, U.; Lueder, G.; Trautner, T.A.; Alonso,
            J.C.

#journal    Gene (1993) 129:41-49
#title      Sequence analysis of the left end of the Bacillus subtilis
            bacteriophage SPp1 genome.
#cross-references MUID:93328123
#accession  JN0741
#molecule_type DNA
#residues   1-60 #label CHA
#cross-references EMBL:X65941; NID:914843; PID:g579088
SUMMARY     #length 60 #molecular-weight 6692 #checksum 7702

Query Match      63.4%; Score 52; DB 2; Length 60;
Best Local Similarity 50.0%; Pred. No. 1.02e+01;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 28 YNGYIKTGF 37
      :|::|:
QY 8 FGQGYVQTFF 17

RESULT 14
ENTRY
TITLE      A24994 #type complete
            cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) A precursor -
            Cellulomonas fimi
ALTERNATE_NAMES
            exo-cellobiohydrolase A
ORGANISM   #formal_name Cellulomonas fimi
DATE       15-Dec-1988 #sequence_revision 22-Nov-1996 #text_change
            10-Jul-1998
ACCESSIONS A24994
REFERENCE   A24994
#authors    O'Neill, G.; Goh, S.H.; Warren, R.A.J.; Kilburn, D.G.; Miller
            Jr., R.C.

#journal    Gene (1986) 44:325-330
#title      Structure of the gene encoding the exoglucanase of
            Cellulomonas fimi.
#cross-references MUID:87055250
#accession  A24994
#molecule_type DNA
#residues   1-484 #label ONE
#cross-references GB:M15824; NID:g144424; PID:g144425

```

##note the amino-terminal sequence of the mature protein (residues 42-71) has been determined (M.L. Langsford, unpublished data)

REFERENCE A67217 White, A.; Withers, S.G.; Gilkes, N.R.; Rose, D.R.  
#authors Submitted to the Brookhaven Protein Data Bank, July 1994  
#submission  
#cross-references PDB:2EXO  
#contents annotation: X-ray crystallography, 1.8 angstroms, residues 42-173, 'G', 175, 'RR', 178-353

REFERENCE A55905 White, A.; Withers, S.G.; Gilkes, N.R.; Rose, D.R.  
#authors Biochemistry (1994) 33:12546-12552  
#journal Crystal structure of the catalytic domain of the beta-1,4-glycanase Cex from Cellulomonas fimi.  
#title  
#contents annotation: X-ray crystallography, 1.8 angstroms, residues 42-173, 'G', 175, 'RR', 178-353

## GENETICS

#gene cex

#description hydrolyses 1,4-beta-D-glucosidic linkages in cellulose and, releasing cellobiose from the non-reducing ends of the chains

CLASSIFICATION #superfamily Cellulomonas cellulose 1,4-beta-cellobiosidase  
A: bacterial cellulose-binding domain homology;  
Streptomyces endo-1,4-beta-xylanase A homology;  
KEYWORDS extracellular protein; glycoprotein; glycosidase; hydrolase; polysaccharide degradation; tandem repeat

## FEATURE

1-41 #domain signal sequence #status predicted #label SIG\  
42-484 #product cellulose 1,4-beta-cellobiosidase #status experimental #label MAT\  
73-353 #domain streptomyces endo-1,4-beta-xylanase A homology #label SXV\  
357-377 #region 7-residue repeats (P-T-P-T-P-T-[S/T])\  
381-482 #domain bacterial cellulose-binding domain homology #label BCB\  
137,398,403,447, #binding\_site carbohydrate (Asn) (covalent) #status predicted\  
461168,274 #active\_site Glu #status experimental\  
208-240,302-308 #disulfide\_bonds #status experimental\  
382-481 #disulfide\_bonds #status predicted  
SUMMARY #length 484 #molecular-weight 51291 #checksum 9660Query Match 63.4%; Score 52; DB 1; Length 484;  
Best Local Similarity 50.0%; Pred. No. 1.02e+01;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 186 LGNGYIETAF 195

:| |||:|  
Qy 8 FGQGYVQTPE 17

## RESULT 15

ENTRY A44766 #type complete  
TITLE defective chorion-1 fci25 protein precursor - fruit fly (Drosophila melanogaster)  
ORGANISM #formal\_name Drosophila melanogaster  
DATE 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 24-Sep-1998

ACCESSIONS A44766

REFERENCE A44766

#authors Waring, G.L.; Hawley, R.J.; Schoenfeld, T.  
#journal Dev. Biol. (1990) 142:1-12  
#title Multiple proteins are produced from the dec-1 eggshell gene in Drosophila by alternative RNA splicing and proteolytic cleavage events.

#cross-references MUID:91032553

#accession A44766

#status preliminary

#molecule\_type mRNA

#residues 1-123 #label WAR

#cross-references GB:M35887; NID:gl57181; PID:gl57182

## GENETICS

#gene FlyBase:dec-1

SUMMARY ##cross-references FlyBase:FBgn0000427  
#length 1123 #molecular-weight 127959 #checksum 6533Query Match 63.4%; Score 52; DB 2; Length 1123;  
Best Local Similarity 50.0%; Pred. No. 1.02e+01;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1098 FIOGYAESPY 1107

:| |||:|

Qy 8 FGQGYVQTPE 17

Search completed: Wed Sep 1 16:22:40 1999  
Job time : 16 secs.

\*\*\*\*\*  
W A P S R E H (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:20:53 1999; MasPar time 3.79 Seconds  
Tabular output not generated. 74.682 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (8-17) from PCTUS9913024.pep (8 of 12)  
Perfect Score: 82  
Sequence: 1 FGQGYVQTFF 10

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 24.996; Variance 29.785; scale 0.839

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	57	69.5	262	1 YG56_YEAST	HYPOTHETICAL 29.4 KD P	3.49e+01
2	55	67.1	1036	1 NIT2_NEUCR	NITROGEN CATABOLIC ENZ	9.37e+01
3	54	65.9	547	1 HBPA_HAEIN	HEME-BINDING PROTEIN A	1.52e+00
4	53	64.6	359	1 OM21_HAEIN	OUTER MEMBRANE PROTEIN	2.46e+00
5	53	64.6	361	1 OM22_HAEIN	OUTER MEMBRANE PROTEIN	2.46e+00
6	53	64.6	371	1 OM25_HAEIN	OUTER MEMBRANE PROTEIN	2.46e+00
7	53	64.6	392	1 PATA_BACSU	PUTATIVE AMINOTRANSFER	2.46e+00
8	52	63.4	488	1 GUX_CELFI	EXOGUCANASE PRECURSOR	3.94e+00
9	52	63.4	1123	1 DC11_DROME	DEFECTIVE CHORION-1 FC	3.94e+00
10	51	62.2	415	1 RL3_DROME	60S RIBOSOMAL PROTEIN	6.29e+00
11	51	62.2	743	1 YCR6_YEAST	HYPOTHETICAL 84.9 KD P	6.29e+00
12	51	62.2	830	1 RP0D_EUGGR	DNA-DIRECTED RNA POLYM	6.29e+00
13	50	61.0	488	1 LYSF_SCOLI	LYSINE-SPECIFIC PERMEA	9.95e+00
14	49	59.8	115	1 TR10_HUMAN	THYROID RECEPTOR INTER	1.56e+01
15	49	59.8	381	1 DBB2_MOUSE	ESTRADIOL 17 BETA-DEHY	1.56e+01
16	49	59.8	699	1 NINA_EMENI	G2-SPECIFIC PROTEIN KI	1.56e+01
17	49	59.8	722	1 GUNF_CLOCE	ENDOGUCANASE F PRECUR	1.56e+01
18	49	59.8	779	1 NIM1_NEUCR	G2-SPECIFIC PROTEIN KI	1.56e+01
19	49	59.8	786	1 AAC_ACTUT	ACULEAGIN A ACYLASE PR	1.56e+01
20	49	59.8	930	1 DP01_HAEIN	DNA POLYMERASE I (EC 2	1.56e+01
21	49	59.8	942	1 DP01_CHLAU	DNA POLYMERASE I (EC 2	1.56e+01
22	48	58.5	381	1 DHB2_RAT	ESTRADIOL 17 BETA-DEHY	2.44e+01
23	48	58.5	402	1 Y718_WETJA	HYPOTHETICAL PROTEIN M	2.44e+01

24	48	58.5	481	1 BIND_STRPU	BINDIN PRECURSOR.	2.44e+01
25	48	58.5	1025	1 ITA8_HUMAN	INTEGRIN ALPHA-8.	2.44e+01
26	47	57.3	57	1 PRPB_HUMAN	PROLINE-RICH PEPTIDE P	3.77e+01
27	47	57.3	82	1 YK2_SSV1	HYPOTHETICAL 9.2 KD PR	3.77e+01
28	47	57.3	164	1 VE18_NPVAC	EARLY 18.5 KD PROTEIN.	3.77e+01
29	47	57.3	166	1 INC_PIG	INTERFERON GAMMA PRECU	3.77e+01
30	47	57.3	177	1 FABA_HAEIN	3-HYDROXYDECANOYL- (AC	3.77e+01
31	47	57.3	229	1 YXDJ_BACSU	HYPOTHETICAL 26.6 KD S	3.77e+01
32	47	57.3	233	1 YP11_VIBAL	HYPOTHETICAL PROTEIN I	3.77e+01
33	47	57.3	234	1 YGGS_SCOLI	HYPOTHETICAL 25.8 KD P	3.77e+01
34	47	57.3	237	1 YGGS_HAEIN	HYPOTHETICAL PROTEIN H	3.77e+01
35	47	57.3	239	1 YQCB_HAEIN	HYPOTHETICAL 37.0 KD P	3.77e+01
36	47	57.3	244	1 Y4SG_CAEEL	HYPOTHETICAL 27.2 KD P	3.77e+01
37	47	57.3	336	1 Y4SG_RHISN	HYPOTHETICAL 37.0 KD P	3.77e+01
38	47	57.3	400	1 GCST_YEAST	AMINOMETHYLTRANSFERASE	3.77e+01
39	47	57.3	462	1 UN47_CAEEL	UNC-47 PROTEIN.	3.77e+01
40	47	57.3	528	1 RPOB_ASTLO	DNA-DIRECTED RNA POLYM	3.77e+01
41	47	57.3	600	1 XOXF_PARDE	PUTATIVE DEHYDROGENASE	3.77e+01
42	47	57.3	602	1 HMD2_LYCES	3-HYDROXY-3-METHYLGUT	3.77e+01
43	47	57.3	646	1 NODQ_RHIS3	PROBABLE SULFATE ADENY	3.77e+01
44	47	57.3	913	1 EDD1_HUMAN	EPITHELIAL DISCOIDIN D	3.77e+01
45	47	57.3	1333	1 YN99_YEAST	PROBABLE ATP-DEPENDENT	3.77e+01

ALIGNMENTS

RESULT 1  
ID YG56\_YEAST STANDARD; PRT; 262 AA.  
AC P53179;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 29.4 KD PROTEIN IN SUGI-RNA15 INTERGENIC REGION.  
GN YG1045W.  
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;  
OC SACCCHAROMYCETACEAE; SACCCHAROMYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C;  
RX MEDLINE; 97377993.  
RA FEUERHANN M., DE MONTIGNY J., POTIER S., SOUCIET J.-L.;  
RT \*The characterization of two new clusters of duplicated genes  
RT suggests a 'Lego' organization of the yeast Saccharomyces cerevisiae  
RT chromosomes\*.  
RL YEAST 13:861-869(1997).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL: Z72568; E243943; -  
CC HYPOTHETICAL PROTEIN.  
KW SEQUENCE 262 AA; 29440 MW; DCD9CB26 CRC32;  
SQ

Query Match 69.5%; Score 57; DB 1; Length 262;  
Best Local Similarity 60.0%; Pred. No. 3.49e+01;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db	107 YGQYVKTAF 116	
Qy	8 FGQGYVQTFF 17	
RESULT 2		
ID NIT2_NEUCR STANDARD; PRT; 1036 AA.		
AC P19212;		
DT 01-NOV-1990 (REL. 16, CREATED)		
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)		



```

Query Match      65.9%; Score 54; DB 1; Length 547;
Best Local Similarity 85.7%; Pred. No. 1.52e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 528 GYVQSPF 534
      |||||
QY 11 GYVOTPF 17

RESULT 4
ID OM21_HAEIN STANDARD; PRT; 359 AA.
AC P43839;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2 OR H10139.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20;
RX MEDLINE; 95350630.
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA VENTER J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL SCIENCE 269:496-512(1995).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: TO PHOTOBACTERIUM STRAIN SS9 OMPH.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32699; G1573092; -
CC TIGR: H10139; -
CC OUTER MEMBRANE; TRANSMEMBRANE; PORIN; SIGNAL.
KW SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 359 OUTER MEMBRANE PROTEIN P2.
FT CHAIN 21 359 OUTER MEMBRANE PROTEIN P2.
SQ SEQUENCE 359 AA; 39375 MW; 40615FB6 CRC32;

Query Match      64.6%; Score 53; DB 1; Length 359;
Best Local Similarity 50.0%; Pred. No. 2.46e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 YAGGYLETFR 87
      :|||::||
QY 8 FGQGYVOTPF 17

RESULT 5
ID OM22_HAEIN STANDARD; PRT; 361 AA.
AC P20149;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2.
OS HAEMOPHILUS INFLUENZAE.

Query Match      64.6%; Score 53; DB 1; Length 361;
Best Local Similarity 50.0%; Pred. No. 2.46e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 YAGGYLETFR 87
      :|||::||
QY 8 FGQGYVOTPF 17

Query Match      64.6%; Score 53; DB 1; Length 361;
Best Local Similarity 50.0%; Pred. No. 2.46e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 YAGGYLETFR 87
      :|||::||
QY 8 FGQGYVOTPF 17

RESULT 6
ID OM25_HAEIN STANDARD; PRT; 371 AA.
AC P46027;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T1, AND T2;
RX MEDLINE; 94293786.
RA DUIM B., VAN ALPHEN L., EIJK P., JANSEN H.M., DANKERT J.;
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major

```

```

OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE B;
RX MEDLINE; 89079316.
RA MUNSON R.S. JR., TOLAN R.W. JR.;
RT "Molecular cloning, expression, and primary sequence of outer
RT membrane protein P2 of Haemophilus influenzae type b.";
RL INFECT. IMMUN. 57:88-94(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE B;
RX MEDLINE; 90158127.
RA MUNSON R.S. JR., BAILEY C., GRASS S.;
RT "Diversity of the outer membrane protein P2 gene from major clones of
RT Haemophilus influenzae type b.";
RL MOL. MICROBIOL. 3:1797-1803(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE B;
RX MEDLINE; 89173305.
RA HANSEN E.J., HASEMANN C., CLAUSELL A., CAPRA J.D., ORTH K.,
RA MOOMAW C.R., SLAUGHTER C.A., LATIMER J.L., MILLER E.E.;
RT "Primary structure of the porin protein of Haemophilus influenzae
RT type b determined by nucleotide sequence analysis.";
RL INFECT. IMMUN. 57:1100-1107(1989).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: TO PHOTOBACTERIUM STRAIN SS9 OMPH.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J03359; G148959; -
CC PIR: A09003; G412251; -
CC PIR: A30542; A30542.
CC OUTER MEMBRANE; TRANSMEMBRANE; PORIN; SIGNAL.
KW SIGNAL 1 20
FT CHAIN 21 361 OUTER MEMBRANE PROTEIN P2.
SQ SEQUENCE 361 AA; 39701 MW; 291391ED CRC32;

Query Match      64.6%; Score 53; DB 1; Length 361;
Best Local Similarity 50.0%; Pred. No. 2.46e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 YAGGYLETFR 87
      :|||::||
QY 8 FGQGYVOTPF 17

RESULT 6
ID OM25_HAEIN STANDARD; PRT; 371 AA.
AC P46027;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T1, AND T2;
RX MEDLINE; 94293786.
RA DUIM B., VAN ALPHEN L., EIJK P., JANSEN H.M., DANKERT J.;
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major

```



```

FT  DISULFID 208 240
FT  DISULFID 302 308
FT  DISULFID 382 481
FT  MUTAGEN 168 168
SQ  SEQUENCE 484 AA; 51291 MW; 2624B193 CRC32;
E->A,D,G: REDUCED ACTIVITY.

Query Match 63.4%; Score 52; DB 1; Length 484;
Best Local Similarity 50.0%; Pred. No. 3,94e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 186 LGNGYIETAF 195
QY 8 FGQGYVQTFP 17
: | | | : | |
: | | | : | |

RESULT 9
ID DC11_DROME STANDARD; PRT; 1123 AA.
AC P18169;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE DEFECTIVE CHORION-1 FC125 PROTEIN PRECURSOR.
GN DEC-1.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
[1]
SEQUENCE FROM N.A.
RX MEDLINE; 91032553.
RT WARKING G.L., HAWLEY R.J., SCHOENFELD T.;
RA "Multiple proteins are produced from the dec-1 eggshell gene in
RT Drosophila by alternative RNA splicing and proteolytic cleavage
events";
RL DEV. BIOL. 142:1-12(1990).
CC -1- FUNCTION: THE DEC-1 FEMALE-STERILE LOCUS PROTEIN IS REQUIRED
CC FOR PROPER ASSEMBLY OF THE DROSOPHILA EGGSHELL.
CC -1- ALTERNATIVE PRODUCTS: MULTIPLE PROTEINS ARE PRODUCED FROM THE
CC DEC-1 GENE BY ALTERNATIVE RNA SPLICING AND PROTEOLYTIC CLEAVAGE
CC EVENTS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M35887; G157182; -
CC PIR; A44766; A44766.
CC FLYBASE; FBgn0000427; dec-1.
CC CHORION; EGGSHELL; REPEAT; ALTERNATIVE SPLICING; SIGNAL.
FT SIGNAL 1 23
FT CHAIN 24 1123
FT DOMAIN 493 788
SQ SEQUENCE 1123 AA; 127959 MW; 99218239 CRC32;
12 X 26 AA TANDEM REPEATS.
DEFECTIVE CHORION-1 FC125 PROTEIN.

Query Match 63.4%; Score 52; DB 1; Length 1123;
Best Local Similarity 50.0%; Pred. No. 3,94e+00;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1098 FIOGYAESPY 1107
QY 8 FGQGYVQTFP 17
: | | | : | |
: | | | : | |

RESULT 10
ID RL3_DROME STANDARD; PRT; 415 AA.
AC O16797;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

```

DE 60S RIBOSOMAL PROTEIN L3.  
GN RPL3  
OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;  
OC DROSOPHILIDAE; DROSOPHILA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHAN H.Y.E., ZHANG Y., HOHEISEL J.D., O'KANE C.J.;  
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -- FUNCTION: THE L3 PROTEIN IS A COMPONENT OF THE LARGE SUBUNIT OF  
CC CYTOPLASMIC RIBOSOMES.  
CC -- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -- SIMILARITY: BELONGS TO THE L3P FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AF016835; G2384754; -  
CC FLYBASE: FBgn0020910; Rpl3.  
CC PROSITE: PS00474; RIBOSOMAL\_L3; 1.  
CC PFAM: PF00297; L3; 1.  
CC RIBOSOMAL PROTEIN.  
CC INIT\_MET 0 BY SIMILARITY.  
CC SEQUENCE 415 AA; 46784 MW; FAD30713 CRC32;  
Query Match 62.2%; Score 51; DB 1; Length 415;  
Best Local Similarity 71.4%; Pred. No. 6.29e+00;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 90 GYIETPF 96  
QY 11 GYVQTFP 17  
II::IIII  
RESULT 11  
ID YCR6\_YEAST STANDARD; PRT: 743 AA.  
AC P25353;  
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)  
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 84.9 KD PROTEIN IN PMP1-PEN2 INTERGENIC REGION.  
GN YCR026C OR YCR26C OR YCR246.  
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;  
OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.  
RN [1]  
RP SEQUENCE OF 1-244 FROM N.A.  
RA POHL F., RICHTERICH P., WURST H.;  
RL SUBMITTED (MAR-1992) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE OF 244-743 FROM N.A.  
RX MEDLINE: 92245758.  
RA BOLLE P.-A., GILLIQUET V., BERBEN G., DUMONT J., HILGER F.;  
RT "The complete sequence of K3B, a 7.9 kb fragment between PGK1 and  
RT CRY1 on chromosome III, reveals the presence of seven open reading  
RT frames.";  
RL YEAST 8:205-213(1992).  
CC -- SIMILARITY: TO YEAST YEL016C.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: X59720; E264490; -  
DR PIR: S19437; S19437.  
DR PIR: S27380; S27380.  
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.  
FT TRANSMEM 114 135 POTENTIAL.  
SQ SEQUENCE 743 AA; 84915 MW; 957CE7AF CRC32;  
Query Match 62.2%; Score 51; DB 1; Length 743;  
Best Local Similarity 60.0%; Pred. No. 6.29e+00;  
Matches 6; Conservative 2; Mismatches 1; Indels 1; Gaps 1;  
Db 563 FPGYIE-PF 571  
QY 8 FCGYVQTFP 17  
II::II::II  
RESULT 12  
ID RPOD\_EUGGR STANDARD; PRT: 830 AA.  
AC P23581;  
DT 01-NOV-1991 (REL. 20, CREATED)  
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE DNA-DIRECTED RNA POLYMERASE BETA\* CHAIN (EC 2.7.7.6).  
GN RPOC2.  
OS EUGLENA GRACILIS.  
OC CHLOROPLAST.  
OC EUKARYOTA; EUGLENOZOA; EUGLENIDA; EUGLENALES; EUGLENA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-Z;  
RX MEDLINE: 90245579.  
RA YEPIZ-PLASCENCIA G.M., RADEBAUGH C.A., HALLICK R.B.;  
RT "The Euglena gracilis chloroplast rpoB gene. Novel gene organization  
RT and transcription of the RNA polymerase subunit operon.";  
RL NUCLEIC ACIDS RES. 18:1869-1878(1990).  
RN [2]  
RP SEQUENCE OF 586-830 FROM N.A.  
RX STRAIN-Z;  
RX MEDLINE: 82142472.  
RA OROZCO E.M., HALLICK R.B.;  
RT "Euglena gracilis chloroplast transfer RNA transcription units. II.  
RT Nucleotide sequence analysis of a tRNAVal-tRNAAsn-tRNAArg-tRNALeu  
RT gene cluster.";  
RL J. BIOL. CHEM. 257:3265-3275(1982).  
CC -- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
CC SUBSTRATES.  
CC -- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +  
CC RNA(N).  
CC -- SUBUNIT: IN CHLOROPLAST THE RNA POLYMERASE IS COMPOSED OF FOUR  
CC SUBUNITS: ALPHA, BETA, BETA', AND BETA".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X17191; G311711; -  
CC EMBL: M22010; G336887; -  
CC EMBL: X70810; G415792; -  
DR PIR: S19259; RNEG82.  
DR PIR: S34554; S34554.  
DR PFAM: PF00623; RNA\_pol\_A; 1.  
DR MENDEL: 5016; EUGgr:RPOC2;1.  
KW TRANSCRIPTION; DNA-DIRECTED RNA POLYMERASE; CHLOROPLAST.  
FT CONFLICT 593 593 I -> II (IN REF. 2).  
FT CONFLICT 688 688 L -> R (IN REF. 2).  
SQ SEQUENCE 830 AA; 94756 MW; 003F633F CRC32;  
Query Match 62.2%; Score 51; DB 1; Length 830;



Best Local Similarity 50.0%; Pred. No. 6.29e+00; Indels 0; Gaps 0;  
Matches 5; Conservative 3; Mismatches 289

Db 280 FPEVIRTPF 289  
QY 8 FGQYVQTF 17

RESULT 13  
ID LYS\_P\_ECOLI STANDARD; PRT; 488 AA.  
AC P25737;  
DT 01-MAY-1992 (REL. 22, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE LYSINE-SPECIFIC PERMEASE.  
GN LYS\_P OR CADR.  
OS ESCHERICHIA COLI.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
OC ESCHERICHIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-BPR2;  
RX MEDLINE; 92250419.  
RA STEFFES C., ELLIS J., WU J., ROSEN B.P.;  
RT "The lysP gene encodes the lysine-specific permease.";  
RL J. BACTERIOL. 174:3242-3249(1992).  
RN [2]  
RP REVISIONS.  
RA ROSEN B.P.;  
RL SUBMITTED (APR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / BHB2600;  
RA RICHTERICH P., LAKEY N., GRYAN G., JAEHN L., MINTZ L., ROBISON K.,  
RA CHURCH G.M.;  
RL SUBMITTED (OCT-1993) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE; 97426617.  
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,  
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GORDEN M.A., ROSE D.J.,  
RA WU B., SHAO Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL SCIENCE 277:1453-1474(1997).  
RN [5]  
RP TOPOLOGY, AND SEQUENCE OF 1-6.  
RX MEDLINE; 96032017.  
RA ELLIS J., CARLIN A., STEFFES C., WU J., LIU J., ROSEN B.P.;  
RT "Topological analysis of the lysine-specific permease of Escherichia coli.";  
RL MICROBIOLOGY 141:1927-1935(1995).  
CC -!- FUNCTION: PERMEASE THAT IS INVOLVED IN THE TRANSPORT ACROSS THE CYTOPLASMIC MEMBRANE OF LYSINE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.  
CC -!- INDUCTION: HIGH, IN ANAEROBIC MEDIA OF LOW PH, CONTAINING LYSINE.  
CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M89774; G466778; -  
DR EMBL; U00007; G405901; -  
DR EMBL; AE000305; G1788480; -  
DR EMBL; S24560; S24560.  
DR ECOCENE; EG11337; LYS\_P.  
DR PROSITE; PS00218; AMINO\_ACID\_PERMEASE; 1.

PFAM; PF00324; aa\_permeases; 1.  
KW TRANSPORT; AMINO-ACID TRANSPORT; TRANSMEMBRANE; INNER MEMBRANE.  
FT INIT MET 0 0  
FT DOMAIN 1 20 CYTOPLASMIC.  
FT TRANSMEM 21 41  
FT DOMAIN 42 54 PERIPLASMIC.  
FT TRANSMEM 55 75  
FT DOMAIN 76 103 CYTOPLASMIC.  
FT TRANSMEM 104 124  
FT DOMAIN 125 129 PERIPLASMIC.  
FT TRANSMEM 130 148  
FT DOMAIN 149 159  
FT TRANSMEM 160 180  
FT DOMAIN 181 198  
FT TRANSMEM 199 217 PERIPLASMIC.  
FT DOMAIN 218 243  
FT TRANSMEM 244 263  
FT DOMAIN 284 282  
FT TRANSMEM 283 303  
FT DOMAIN 304 346  
FT TRANSMEM 347 364  
FT DOMAIN 365 370  
FT TRANSMEM 371 391  
FT DOMAIN 392 414  
FT TRANSMEM 415 432  
FT DOMAIN 433 448  
FT TRANSMEM 449 469  
FT DOMAIN 470 488  
FT CONFLICT 121 121 N -> S (IN REF. 1).  
SQ SEQUENCE 488 AA; 53471 MW; B0E7230B CRC32;

Query Match 61.0%; Score 50; DB 1; Length 488;  
Best Local Similarity 50.0%; Pred. No. 9.95e+00;  
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 84 YGONYVEEGF 93  
QY 8 FGQYVQTF 17

RESULT 14  
ID TR10\_HUMAN STANDARD; PRT; 115 AA.  
AC Q15642;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE THYROID RECEPTOR INTERACTING PROTEIN 10 (TRIP10) (FRAGMENT).  
GN TRIP10.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95295737.  
RA LEE J.W., CHOI H.-S., GYURIS J., BRENT R., MOORE D.D.;  
RT "Two classes of proteins dependent on either the presence or absence of thyroid hormone for interaction with the thyroid hormone receptor.";  
RL MOL. ENDOCRINOL. 9:243-254(1995).  
CC -!- FUNCTION: THYROID RECEPTOR INTERACTING PROTEINS (TRIPS) SPECIFICALLY INTERACT WITH THE LIGAND BINDING DOMAIN OF THE THYROID RECEPTOR (TR). TRIP10 REQUIRES THE PRESENCE OF THYROID HORMONE FOR ITS INTERACTION.  
CC -!- TISSUE SPECIFICITY: EXPRESSED MOST ABUNDANTLY IN SKELETAL MUSCLE, HEART AND LUNG.  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
DR EMBL; L40379; G695376; 1.  
DR PROSITE; PS50002; SH3; 1.  
DR PFAM; PF00018; SH3; 1.  
KW SH3 DOMAIN.  
FT NON\_TER 1 1  
FT DOMAIN 60 115 SH3.  
SQ SEQUENCE 115 AA; 12848 MW; E678DBE7 CRC32;  
  
Query Match 59.8%; Score 49; DB 1; Length 115;  
Best Local Similarity 55.6%; Pred.No. 1.56e+01;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
Db 101 GEGYVPTSY 109  
1:|||||:  
QY 9 GGGYVQTFF 17

RESULT 15  
ID DHB2\_MOUSE STANDARD; PRT; 381 AA.  
AC P51658; O08898;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE ESTRADIOL 17 BETA-DEHYDROGENASE 2 (EC 1.1.1.62) (17-BETA-HSD 2)  
DE (17-BETA-HYDROXYSTEROID DEHYDROGENASE 2).  
GN HSD17B2 OR EDH17B2.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-BALB/C; TISSUE=LIVER;  
RX MEDLINE; 97344259.  
RA MUSTONEN M.;  
RT \*Cloning of mouse 17beta-hydroxysteroid dehydrogenase type 2, and  
RT analysing expression of the mRNAs for types 1, 2, 3, 4 and 5 in mouse  
RL embryos and adult tissues.;  
RL BIOCHEM. J. 325:199-205(1997).  
RN [2]  
RP SEQUENCE OF 1-358 FROM N.A.  
RC STRAIN-BALB/C;  
RA STOFFEL W.; WEISS B.;  
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- FUNCTION: CAPABLE OF CATALYZING THE INTERCONVERSION OF  
CC TESTOSTERONE AND ANDROSTENEDIONE, AS WELL AS ESTRADIOL AND  
CC ESTRONE. ALSO HAS 20-ALPHA-HSD ACTIVITY. USES NADH WHILE EDH17B3  
CC USES NADPH (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: ESTRADIOL-17-BETA + NAD(+) -> ESTRONE +  
CC NADH.  
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
CC FAMILY (SDR).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Y09517; E289833; -.  
DR EMBL; X95685; E223346; -.  
DR MGD; MGI:1096386; HSD17B2.  
DR PROSITE; PS00061; ADH\_SHORT; 1.  
DR PFAM; PF00106; adh\_short; 1.  
DR HSSP; P14061; 1FDV.  
KW STEROID BIOSYNTHESIS; OXIDOREDUCTASE; NAD; MULTIGENE FAMILY;  
KW TRANSMEMBRANE; SIGNAL-ANCHOR.  
FT TRANSMEM 4 24 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
FT NP\_BIND 83 112 NAD (BY SIMILARITY).

FT ACT\_SITE 233 233 BY SIMILARITY.  
FT CONFLICT 36 37 QA -> RP (IN REF. 2).  
SQ SEQUENCE 381 AA; 41835 MW; 6A388BC8 CRC32;  
  
Query Match 59.8%; Score 49; DB 1; Length 381;  
Best Local Similarity 62.5%; Pred.No. 1.56e+01;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
Db 296 YGQDYVHT 303  
:|||||:  
QY 8 FGQGYVOT 15

Search completed: Wed Sep 1 16:21:06 1999  
Job time : 13 secs.

\*\*\*\*\*  
 W A I S R L A  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:21:25 1999; MasPar time 5.97 Seconds

91.381 Million cell updates/sec

Tabular output not generated.

Title: >PCT-US99-13024-2  
 Description: (8-17) from PCTUS9913024.pep (8 of 12)  
 Perfect Score: 82  
 Sequence: 1 FGQGVQTPF 10

Scoring table: PAM 150  
 Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: sprembl9

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
 5:sp-invertebrate 6:sp-mammal 7:sp\_mhc 8:sp-organelle  
 9:sp-phage 10:sp-plant 11:sp-rodent 12:sp\_unclassified  
 13:sp-vertebrate 14:sp\_virus

Statistics: Mean 23.872; Variance 31.947; scale 0.747

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	55	67.1	509	1	028347 LONG-CHAIN-FATTY-ACID-	2.79e+00
2	54	65.9	722	3	013839 PUTATIVE G2-SPECIFIC P	4.38e+00
3	53	64.6	209	5	082250 F3F8.3 PROTEIN.	6.83e+00
4	53	64.6	344	2	048025 OUTER MEMBRANE PROTEIN	6.83e+00
5	53	64.6	357	2	048225 MAJOR OUTER MEMBRANE P	6.83e+00
6	53	64.6	357	2	001452 OUTER MEMBRANE PROTEIN	6.83e+00
7	53	64.6	364	2	068632 MAJOR OUTER MEMBRANE P	6.83e+00
8	53	64.6	364	2	068634 MAJOR OUTER MEMBRANE P	6.83e+00
9	53	64.6	365	2	068635 MAJOR OUTER MEMBRANE P	6.83e+00
10	53	64.6	365	2	068631 MAJOR OUTER MEMBRANE P	6.83e+00
11	53	64.6	368	2	068633 MAJOR OUTER MEMBRANE P	6.83e+00
12	53	64.6	368	2	068630 MAJOR OUTER MEMBRANE P	6.83e+00
13	53	64.6	390	2	001450 MAJOR OUTER MEMBRANE P	6.83e+00
14	53	64.6	584	5	019672 F2IC3.4 PROTEIN.	6.83e+00
15	53	64.6	686	4	060294 KIAA0347 PROTEIN.	6.83e+00
16	53	64.6	936	2	025141 CYTOCHROME C BIOGENESI	6.83e+00
17	52	63.4	60	9	038080 ORE50.	1.06e+01
18	52	63.4	485	2	059277 BETA-1,4-XYLANASE PREC	1.06e+01
19	51	62.2	65	7	P87432 LEPIDODACTYLUS MOESTUS	1.63e+01
20	51	62.2	340	2	067489 HYPOTHETICAL 37.1 KD P	1.63e+01

21	51	62.2	400	4	092734 TGF PROTEIN.	1.63e+01
22	51	62.2	2301	10	080784 F13P17.19 PROTEIN.	1.63e+01
23	51	62.2	052645	2	052545 POLYKETIDE SYNTHASE.	1.63e+01
24	51	62.2	5069	2	052789 RIFAMYCIN POLYKETIDE S	1.63e+01
25	50	61.0	234	2	025327 MEMBRANE FUSION PROTEI	2.51e+01
26	50	61.0	241	2	068961 PUTATIVE MEMBRANE FUSI	2.51e+01
27	50	61.0	566	2	P71093 NEOPULLULANASE.	2.51e+01
28	50	61.0	026441	5	026441 RK2-GLIAL-SPECIFIC HOM	2.51e+01
29	50	61.0	612	5	024477 REVERSED POLARITY.	2.51e+01
30	50	61.0	612	5	024477 LIPASE (EC 3.1.1.3).	2.51e+01
31	50	61.0	613	2	059933 EXTRACELLULAR LIPASE (	2.51e+01
32	50	61.0	792	10	048572 SIMILAR TO MAIZE TRANS	2.51e+01
33	50	61.0	793	5	022047 TO1B7.6 PROTEIN.	2.51e+01
34	50	61.0	803	5	046129 TRP-LIKE PROTEIN.	2.51e+01
35	49	59.8	107	1	P94122 HYPOTHETICAL 12.0 KD P	3.82e+01
36	49	59.8	248	2	P75004 REP PROTEIN.	3.82e+01
37	49	59.8	255	2	050462 U0002M.	3.82e+01
38	49	59.8	347	2	070080 OMFI PROTEIN (FRAGMENT	3.82e+01
39	49	59.8	347	2	069304 OMP1 (FRAGMENT).	3.82e+01
40	49	59.8	360	2	068623 MAJOR OUTER MEMBRANE P	3.82e+01
41	49	59.8	418	2	P95139 HYPOTHETICAL 45.1 KD P	3.82e+01
42	49	59.8	425	1	059244 425AA LONG HYPOTHETICA	3.82e+01
43	49	59.8	433	10	064988 ACETYL COA: BENZYLALCO	3.82e+01
44	49	59.8	545	4	015184 CDC42-INTERACTING PROT	3.82e+01
45	49	59.8	547	11	P97531 SALT-TOLERANT PROTEIN.	3.82e+01

## ALIGNMENTS

RESULT	ID	028347	PRELIMINARY;	PRT;	509 AA.
AC	028347;				
DT	01-JAN-1998	(TREMBLREL. 05, CREATED)			
DT	01-JAN-1998	(TREMBLREL. 05, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	LONG-CHAIN-FATTY-ACID--COA LIGASE (FADD-8).				
GN	AF1932.				
OS	ARCHAEOGLOBUS FULGIDUS.				
OC	ARCHAEA; EURYARCHAEOTA; ARCHAEOGLOBALES; ARCHAEOGLOBACEAE;				
OC	ARCHAEOGLOBUS.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-VC-16 / DSM 4304 / ATCC 49558;				
RX	MEDLINE; 98049343.				
RA	KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,				
RA	KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,				
RA	RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIES N.C.,				
RA	FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,				
RA	KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,				
RA	PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,				
RA	OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTTERBACK T.,				
RA	COTTON M.D., SPRIGGS T., ARTIACH P., KATNE B.P., SYKES S.M.,				
RA	SADON P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,				
RA	MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,				
RA	VENTER J.C.;				
RT	*The complete genome sequence of the hyperthermophilic,				
RT	sulphate-reducing archaeon Archaeoglobus fulgidus.*;				
RL	NATURE 390:364-370(1997).				
DR	EMBL; AB000970; G2648612; -.				
DR	TIGR; AF1932; -.				
DR	PFAM; PF00501; AMP-binding; 1.				
KW	HYPOTHETICAL PROTEIN; LIGASE.				
SQ	SEQUENCE 509 AA; 57035 MW; 6CD28ACD CRC32;				

Query Match 67.1%; Score 55; DB 1; Length 509;  
 Best Local Similarity 60.0%; Pred. No. 2.79e+00;  
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 325 FTGGYGVQPF 334

QY 8 FGQGVQTPF 17

```
RESULT 2
ID O13839 PRELIMINARY; PRT: 722 AA.
AC O13839:
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PUTATIVE G2-SPECIFIC PROTEIN KINASE C19E9.02 (EC 2.7.1.-).
GN SPAC19E9.02.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972:
RA BADCOCK K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: PROTEIN KINASE THAT PLAYS AN IMPORTANT ROLE IN MITOTIC
CC REGULATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF
CC SER/THR-PROTEIN KINASES. BELONGS TO THE NIMA SUBFAMILY.
DR EMBL: Z98975; E339915; -.
DR PFM; PF00069; Pfkinase.1.
KW HYPOTHETICAL PROTEIN; TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE;
KW ATP-BINDING; MITOSIS; NUCLEAR PROTEIN; CELL CYCLE; CELL DIVISION.
FT DOMAIN 4 281 PROTEIN KINASE (BY SIMILARITY).
FT NP_BIND 10 18 ATP (BY SIMILARITY).
FT BINDING 33 33 ATP (BY SIMILARITY).
FT ACT_SITE 151 151 BY SIMILARITY.
FT ACT_SITE 722 AA; 82689 MW; 286A8C12 CRC32;
SQ SEQUENCE 722 AA; 82689 MW; 286A8C12 CRC32;

Query Match 65.9%; Score 54; DB 3; Length 722;
Best Local Similarity 60.0%; Pred. No. 4.38e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 182 FTQSVGTPY 191
QY 8 FGQGYVQTPF 17

RESULT 3
ID O62260 PRELIMINARY; PRT: 209 AA.
AC O62260:
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE F53F8.3 PROTEIN.
GN F53F8.3.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA MATTHEWS L.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., FAVELLO A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., HILLIER L., JIER M., JOHNSTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN K., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL NATURE 368:32-38 (1994).
CC -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
```

```
DR EMBL: Z81547; E1347158; -.
DR PROSITE: PS00518; ZINC_FINGER_C3HC4; 1.
KW ZINC-FINGER.
SQ SEQUENCE 209 AA; 24425 MW; D1978DC1 CRC32;

Query Match 64.6%; Score 53; DB 5; Length 209;
Best Local Similarity 66.7%; Pred. No. 6.83e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 177 GGGYKICPF 185
QY 9 GGGYVQTPF 17

RESULT 4
ID Q48025 PRELIMINARY; PRT: 344 AA.
AC Q48025:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE OUTER MEMBRANE PROTEIN P2 (FRAGMENT).
GN OMEP2.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-12:
RX MEDLINE: 94245383.
RA BELL J., GRASS S., JEANTEUR D., MUNSON R.S.;
RT "Diversity of the P2 protein among nontypeable Haemophilus influenzae
RT isolates."
RL INFECT. IMMUN. 62:2639-2643 (1994).
DR EMBL: U08206; G559998; -.
FT NON_TER 1 1
SQ SEQUENCE 344 AA; 37781 MW; 257B8DC5 CRC32;

Query Match 64.6%; Score 53; DB 2; Length 344;
Best Local Similarity 50.0%; Pred. No. 6.83e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 58 YAGGYLETRF 67
QY 8 FGQGYVQTPF 17

RESULT 5
ID Q48225 PRELIMINARY; PRT: 357 AA.
AC Q48225:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MAJOR OUTER MEMBRANE PROTEIN P2 (FRAGMENT).
GN P2.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-3221B:
RX MEDLINE: 94018553.
RA DUIM B., DANKERT J., JANSEN H., VAN ALPHEN L.;
RT "Genetic analysis of the diversity in outer membrane protein P2 of
RT non-encapsulated Haemophilus influenzae."
RL MICROB. PATHOG. 14:451-462 (1993).
DR EMBL: X73392; G860954; -.
FT NON_TER 357 357
SQ SEQUENCE 357 AA; 39093 MW; 95AD9567 CRC32;

Query Match 64.6%; Score 53; DB 2; Length 357;
Best Local Similarity 50.0%; Pred. No. 6.83e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

Db 78 YAOQYLETRF 87  
::|::|::|  
QY 8 FGQGYVQTF 17

RESULT 6 PRELIMINARY; PRT; 357 AA.  
ID Q01452  
AC Q01452;  
DT 01-NOV-1996 (TREMREL. 01, CREATED)  
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
DE MAJOR OUTER MEMBRANE PROTEIN P2 PRECURSOR.  
GN OMP2  
OS HAEMOPHILUS INFLUENZAE.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;  
OC HAEMOPHILUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 93084371.  
RA SIKEMA D.J., MURPHY T.F.;  
RT "Molecular analysis of the P2 porin protein of nontypeable  
RT Haemophilus influenzae.";  
RL INFECT. IMMUN. 60:5204-5211(1992).  
DR EMBL: M93270; G148967; -  
KW OUTER MEMBRANE; PORIN; SIGNAL.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 357 OUTER MEMBRANE PROTEIN P2.  
SQ SEQUENCE 357 AA; 39215 MW; C56B40D CRC32;

Query Match 64.6%; Score 53; DB 2; Length 357;  
Best Local Similarity 50.0%; Pred. No. 6.83e+00;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 YAOQYLETRF 87  
::|::|::|  
QY 8 FGQGYVQTF 17

RESULT 7 PRELIMINARY; PRT; 364 AA.  
ID O68632  
AC O68632;  
DT 01-AUG-1998 (TREMREL. 07, CREATED)  
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
DE MAJOR OUTER MEMBRANE PORIN P2 VARIANT (FRAGMENT).  
OS HAEMOPHILUS INFLUENZAE.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;  
OC HAEMOPHILUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-30-77C;  
RX MEDLINE: 94293786.  
RA DUIM B., VAN ALPHEN L., EIJK P., JANSSEN H.M., DANKERT J.;  
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major  
RT outer membrane protein P2 in patients with chronic bronchitis is  
RT caused by point mutations.";  
RL MOL. MICROBIOL. 11:1181-1189(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-30-77C;  
RA REGELINK A.G., DAHAN D., MOLLER L.V.M., COULTON J.W., EIJK P.,  
RA VAN ULSEN P., DANKERT J., VAN ALPHEN L.;  
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AF052550; G2981123; -  
FT NON\_TER 1 1  
FT NON\_TER 364 364  
SQ SEQUENCE 364 AA; 40097 MW; 2D97185B CRC32;

Query Match 64.6%; Score 53; DB 2; Length 364;  
Best Local Similarity 50.0%; Pred. No. 6.83e+00;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 71 YAOQYLETRF 80

QY 8 FGQGYVQTF 17  
::|::|::|

RESULT 8 PRELIMINARY; PRT; 364 AA.  
ID O68634  
AC O68634;  
DT 01-AUG-1998 (TREMREL. 07, CREATED)  
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
DE MAJOR OUTER MEMBRANE PORIN P2 VARIANT (FRAGMENT).  
OS HAEMOPHILUS INFLUENZAE.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;  
OC HAEMOPHILUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-32-82E;  
RX MEDLINE: 94293786.  
RA DUIM B., VAN ALPHEN L., EIJK P., JANSSEN H.M., DANKERT J.;  
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major  
RT outer membrane protein P2 in patients with chronic bronchitis is  
RT caused by point mutations.";  
RL MOL. MICROBIOL. 11:1181-1189(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-32-82E;  
RA REGELINK A.G., DAHAN D., MOLLER L.V.M., COULTON J.W., EIJK P.,  
RA VAN ULSEN P., DANKERT J., VAN ALPHEN L.;  
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AF052552; G2981127; -  
FT NON\_TER 1 1  
FT NON\_TER 364 364  
SQ SEQUENCE 364 AA; 39839 MW; 147EB78E CRC32;

Query Match 64.6%; Score 53; DB 2; Length 364;  
Best Local Similarity 50.0%; Pred. No. 6.83e+00;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 71 YAOQYLETRF 80  
::|::|::|  
QY 8 FGQGYVQTF 17

RESULT 9 PRELIMINARY; PRT; 365 AA.  
ID O68635  
AC O68635;  
DT 01-AUG-1998 (TREMREL. 07, CREATED)  
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
DE MAJOR OUTER MEMBRANE PORIN P2 VARIANT (FRAGMENT).  
OS HAEMOPHILUS INFLUENZAE.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;  
OC HAEMOPHILUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-32-82F;  
RX MEDLINE: 94293786.  
RA DUIM B., VAN ALPHEN L., EIJK P., JANSSEN H.M., DANKERT J.;  
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major  
RT outer membrane protein P2 in patients with chronic bronchitis is  
RT caused by point mutations.";  
RL MOL. MICROBIOL. 11:1181-1189(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-32-82F;  
RA REGELINK A.G., DAHAN D., MOLLER L.V.M., COULTON J.W., EIJK P.,  
RA VAN ULSEN P., DANKERT J., VAN ALPHEN L.;  
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AF052553; G2981129; -  
FT NON\_TER 1 1  
FT NON\_TER 365 365  
SQ SEQUENCE 365 AA; 39953 MW; 3567EC05 CRC32;

Query Match 64.6%; Score 53; DB 2; Length 365;  
Best Local Similarity 50.0%; Pred. No. 6.83e+00;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 71 YAQGYLETRF 80  
QY 8 FGQGYVQTPF 17

RESULT 10  
ID O68631 PRELIMINARY; PRT; 365 AA.  
AC O68631; 1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE MAJOR OUTER MEMBRANE PORIN P2 VARIANT (FRAGMENT).  
OS HAEMOPHILUS INFLUENZAE.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;  
OC HAEMOPHILUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=30-77B;  
RX MEDLINE; 94293786.  
RA DUIM B., VAN ALPHEN L., EIJK P., JANSEN H.M., DANKERT J.;  
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major  
outer membrane protein P2 in patients with chronic bronchitis is  
caused by point mutations.";  
RL MOL. MICROBIOL. 11:1181-1189(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=30-77B;  
RA REGELINK A.G., DAHAN D., MOLLER L.V.M., COULTON J.W., EIJK P.,  
RA VAN ULSEN P., DANKERT J., VAN ALPHEN L.;  
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; AF052549; G2981121; -.  
FT NON\_TER 365  
SQ SEQUENCE 365 AA; 40098 MW; 4ECA46B6 CRC32;  
Query Match 64.6%; Score 53; DB 2; Length 365;  
Best Local Similarity 50.0%; Pred. No. 6.83e+00;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 71 YAQGYLETRF 80  
QY 8 FGQGYVQTPF 17

RESULT 11  
ID O68633 PRELIMINARY; PRT; 368 AA.  
AC O68633; 1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE MAJOR OUTER MEMBRANE PORIN P2 VARIANT (FRAGMENT).  
OS HAEMOPHILUS INFLUENZAE.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;  
OC HAEMOPHILUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=30-77F;  
RX MEDLINE; 94293786.  
RA DUIM B., VAN ALPHEN L., EIJK P., JANSEN H.M., DANKERT J.;  
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major  
outer membrane protein P2 in patients with chronic bronchitis is  
caused by point mutations.";  
RL MOL. MICROBIOL. 11:1181-1189(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=30-77F;  
RA REGELINK A.G., DAHAN D., MOLLER L.V.M., COULTON J.W., EIJK P.,  
RA VAN ULSEN P., DANKERT J., VAN ALPHEN L.;  
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; AF052549; G2981121; -.  
FT NON\_TER 365  
SQ SEQUENCE 365 AA; 40098 MW; 4ECA46B6 CRC32;  
Query Match 64.6%; Score 53; DB 2; Length 365;  
Best Local Similarity 50.0%; Pred. No. 6.83e+00;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

DR EMBL; AF052551; G2981125; -.  
FT NON\_TER 368  
SQ SEQUENCE 368 AA; 40510 MW; 37391D02 CRC32;  
Query Match 64.6%; Score 53; DB 2; Length 368;  
Best Local Similarity 50.0%; Pred. No. 6.83e+00;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 71 YAQGYLETRF 80  
QY 8 FGQGYVQTPF 17

RESULT 12  
ID O68630 PRELIMINARY; PRT; 368 AA.  
AC O68630; 1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE MAJOR OUTER MEMBRANE PORIN P2 VARIANT (FRAGMENT).  
OS HAEMOPHILUS INFLUENZAE.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;  
OC HAEMOPHILUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=30-77A;  
RX MEDLINE; 94293786.  
RA DUIM B., VAN ALPHEN L., EIJK P., JANSEN H.M., DANKERT J.;  
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major  
outer membrane protein P2 in patients with chronic bronchitis is  
caused by point mutations.";  
RL MOL. MICROBIOL. 11:1181-1189(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=30-77A;  
RA REGELINK A.G., DAHAN D., MOLLER L.V.M., COULTON J.W., EIJK P.,  
RA VAN ULSEN P., DANKERT J., VAN ALPHEN L.;  
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; AF052548; G2981119; -.  
FT NON\_TER 368  
SQ SEQUENCE 368 AA; 40556 MW; 14B756EE CRC32;  
Query Match 64.6%; Score 53; DB 2; Length 368;  
Best Local Similarity 50.0%; Pred. No. 6.83e+00;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 71 YAQGYLETRF 80  
QY 8 FGQGYVQTPF 17

RESULT 13  
ID Q01450 PRELIMINARY; PRT; 390 AA.  
AC Q01450; 1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR.  
GN OMPP2.  
OS HAEMOPHILUS INFLUENZAE.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;  
OC HAEMOPHILUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 93084371.  
RA SIKKEMA D.J., MURPHY T.F.;  
RT "Molecular analysis of the P2 porin protein of nontypeable  
Haemophilus influenzae.";  
RL INFECT. IMMUN. 60:5204-5211(1992).  
DR EMBL; M93268; G148963; -.  
PFAM; PF00267; Gram-ve\_porins; 1.

DR EMBL; M93268; G148963; -.  
PFAM; PF00267; Gram-ve\_porins; 1.

```

KW  OUTER MEMBRANE; PORIN; SIGNAL.
FT  SIGNAL 1 20 POTENTIAL.
FT  CHAIN 21 390 OUTER MEMBRANE PROTEIN P2.
SQ  SEQUENCE 390 AA; 42941 MW; 0E424362 CRC32;

Query Match 64.6%; Score 53; DB 2; Length 390;
Best Local Similarity 50.0%; Pred. No. 6.83e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 YAOGYLETRF 87
   ::::|:|
QY 8 FGQGVVQTPF 17

RESULT 14
ID Q19672 PRELIMINARY; PRT; 584 AA.
AC Q19672;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE F2IC3.4 PROTEIN.
GN F2IC3.4.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA MCMURRAY A.;
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SHALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
DR EMBL; Z1261; E1345956;
SQ SEQUENCE 584 AA; 66338 MW; 10A646D4 CRC32;

Query Match 64.6%; Score 53; DB 5; Length 584;
Best Local Similarity 87.5%; Pred. No. 6.83e+00;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 243 GNGYVQTP 250
   |:::|
QY 9 GQGVVQTP 16

RESULT 15
ID O60294 PRELIMINARY; PRT; 686 AA.
AC O60294;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE KIAA0547 PROTEIN.
GN KIAA0547.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-BRAIN.
RX MEDLINE; 98290545.
RA NAGASE T., ISHIKAWA K., MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N.,

```

```

RA OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA RES. 5:31-39(1998).
DR EMBL; AB011119; D1026403;
SQ SEQUENCE 686 AA; 75583 MW; B5C954CA CRC32;

Query Match 64.6%; Score 53; DB 4; Length 686;
Best Local Similarity 66.7%; Pred. No. 6.83e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 29 ARGVVDPPF 37
   ::::|:|
QY 9 GQGVVQTPF 17

Search completed: Wed Sep 1 16:22:04 1999
Job time : 39 secs.

```

\*\*\*\*\*  
 W P S R L  
 \*\*\*\*\* (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:25:40 1999; MasPar time 4.56 Seconds  
 Tabular output not generated. 172.419 Million cell updates/sec

Title: >PCT-US99-13024-2  
 Description: (57-93) from PCTUS9913024.pap (9 of 12)  
 Sequence: 1 SAGLHLRVDPSTGALVDSKYAFSTNSDNTTSAAFVS 37  
 Scoring table: PAM 150  
 Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21 22:part22 23:part23  
 24:part24 25:part25 26:part26 27:part27 28:part28  
 29:part29 30:part30 31:part31 32:part32 33:part33  
 34:part34 35:part35 36:part36 37:part37 38:part38  
 39:part39

Statistics: Mean 22.786; Variance 86.998; scale 0.262

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	73	31.5	272	2	R07522 Alpha-Trichosanthin e	2.30e+01
2	73	31.5	272	5	R25577 Ribosome inactivating	2.30e+01
3	73	31.5	345	10	R52824 GTP-cyclohydrolase II	2.30e+01
4	70	30.2	820	13	R71802 N-acetylheparosan fra	4.14e+01
5	68	29.3	119	34	W77531 Nickel-binding peripl	6.11e+01
6	68	29.3	129	37	W72935 Mycobacterium tubercu	6.11e+01
7	68	29.3	1252	14	R80530 B. sphaericus SLP	6.11e+01
8	67	28.9	371	29	W41513 N. meningitidis alpha	7.42e+01
9	66	28.4	248	23	W21704 Luffin-A	8.99e+01
10	66	28.4	248	7	R37294 Plant type I RIP Luff	8.99e+01
11	66	28.4	248	14	R14180 Type I ribosome-inact	8.99e+01
12	66	28.4	248	3	R12468 Luffa cylindrica bica	8.99e+01
13	66	28.4	248	26	R63906 Type I ribosome-inact	8.99e+01
14	66	28.4	248	24	W25141 Luffin-A (a ribosome	8.99e+01
15	66	28.4	275	10	R53731 Luffin (ribosome inac	8.99e+01
16	66	28.4	277	6	R29909 Prod. of the luffin-f	8.99e+01

17	66	28.4	278	6	R29910	Prod. of the luffin-g	8.99e+01
18	66	28.4	539	16	R79580	Mouse CcT-delta subun	8.99e+01
19	66	28.4	577	22	W20634	H. pylori cytoplasmic	8.99e+01
20	65	28.0	218	23	W15104	Myxoma virus MA56 gen	1.09e+02
21	65	28.0	300	1	R05109	Sequence encoded by R	1.09e+02
22	65	28.0	571	30	W55997	Protein SEQ ID NO:227	1.09e+02
23	65	28.0	811	17	R95014	C. difficile toxin A	1.09e+02
24	65	28.0	812	17	R95017	C. difficile toxin A	1.09e+02
25	65	28.0	970	1	R93305	Antigenic surface pro	1.09e+02
26	65	28.0	970	16	R85479	Japanese Encephalitis	1.09e+02
27	65	28.0	1295	2	R10424	Recombinant vaccinia	1.09e+02
28	65	28.0	1295	1	R05597	Sequence containing J	1.09e+02
29	65	28.0	2431	5	R25138	SFV4 non-structural p	1.09e+02
30	65	28.0	2710	35	W68387	Clostridium difficile	1.09e+02
31	65	28.0	2710	17	R95016	C. difficile toxin A	1.09e+02
32	64	27.6	198	2	R07519	Trichosanthin-like pr	1.32e+02
33	64	27.6	198	5	R25574	Trichosanthin kirilow	1.32e+02
34	64	27.6	251	20	R98392	Maleate isomerase.	1.32e+02
35	64	27.6	251	20	R98391	Maleate isomerase.	1.32e+02
36	64	27.6	432	31	W40594	E. coli ADSS protein.	1.32e+02
37	63	27.2	69	28	W46470	Amino acid sequence o	1.59e+02
38	63	27.2	143	3	P50063	Soybean heat shock pr	1.59e+02
39	63	27.2	1291	35	W59912	Amino acid sequence o	1.59e+02
40	63	27.2	1453	6	R31038	E2 protein of Canine	1.59e+02
41	62	26.7	183	22	W20641	H. pylori cytoplasmic	1.92e+02
42	62	26.7	500	2	P71739	Flavivirus V3 antigen	1.92e+02
43	62	26.7	521	3	P51009	Sequence of neutral p	1.92e+02
44	62	26.7	524	25	W32096	Miniature swine retro	1.92e+02
45	62	26.7	1451	5	R27819	CCVinsavc spike prote	1.92e+02

#### ALIGNMENTS

RESULT 1  
 ID R07522 standard; protein; 272 AA.

AC R07522;  
 DT 06-FEB-1991 (first entry)  
 DE Alpha-Trichosanthin encoded by insert sequence from clone 3.  
 KW trichosanthin [TCS]; Human Immunodeficiency Virus; Inhibition.  
 OS Trichosanthin Kirilowil.  
 FH Key Location/Qualifiers  
 FT protein 24..272  
 FT region /label=alpha-TCS  
 FT region 1..23 /label=signal peptide  
 PN W09012097-A.  
 PD 18-OCT-1990.  
 PF 04-APR-1990; U01816.  
 PR 04-APR-1989; US-333184.  
 PA (GENE-) Genelabs Inc.  
 PI Platek M, Chow T, Fry K;  
 DR WPI; 90-334847/44.  
 DR NP-PSDB; Q06350.  
 PT Recombinant trichosanthin protein - with selective inhibitory effect  
 PT on viral expression in HIV infected T-cells or monocytes/macrophage.  
 PS Example; Fig 19; 102pp; English.  
 CC p030E was used as a probe to identify clones containing sequences  
 CC coding for TCS in a 1.kirilowil genomic library. Five clones were  
 CC eventually isolated and sequenced, including clone 3.  
 CC They were found to have homology to the alpha-TCS coding sequence.  
 CC See also Q06343-9 and Q06351.  
 SQ Sequence 272 AA;

Query Match 31.5%; Score 73; DB 2; Length 272;  
 Best Local Similarity 39.1%; Pred. No. 2.30e+01;  
 Matches 9; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

Db 97 gylvnttsyffnepdgaataskfv 119

QY 70 GALVDSKSYAFSTNSDNTTSAAFV 92

RESULT 2



```

CC amino-2,4-(1H,3H)-pyrimidinol);
CC - rib-4 gene: 6,7-dimethyl-8-ribityllumazine (DMRL)-synthase;
CC - rib-5 gene: riboflavin-synthase;
CC - rib-3 gene: L-3,4-dihydroxy-2-butanone-4-phosphate (DBP)-synthase.
SQ Sequence 345 AA;

Query Match 31.5%; Score 73; DB 10; Length 345;
Best Local Similarity 50.0%; Pred. No. 2.30e+01;
Matches 14; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

Db 114 glale-fddstgellaskattwdahndt 140
||| ||||| :||| : |||
QY 59 GLHLRVDPSTGALVDSKSYAFSTNDT 86

RESULT 4
ID R71802 standard; Protein; 820 AA.
AC R71802;
DT 19-OCT-1995 (first entry)
DE N-acetylheparosan fragmentation enzyme.
KW N-acetylheparosan; polysaccharide K5; fragmentation; degradation;
KW pharmaceutical starting material.
OS Escherichia coli (K5) SEBR 3282 (NCMC I-1013).
FH Key Location/Qualifiers
FT binding_site 384..387
FT /note= "Asn-X1-X2-Ser motif (X1-Ile, Leu or Tyr
FT and X2-Val, Ile or Ala)"
FT binding_site 411..414
FT /note= "Asn-X1-X2-Ser motif (X1-Ile, Leu or Tyr
FT and X2-Val, Ile or Ala)"
FT binding_site 433..436
FT /note= "Asn-X1-X2-Ser motif (X1-Ile, Leu or Tyr
FT and X2-Val, Ile or Ala)"
FT binding_site 461..464
FT /label= potential_lipoprotein_attachment_site
FT /note= "Asn-X1-X2-Ser motif (X1-Ile, Leu or Tyr
FT and X2-Val, Ile or Ala)"
FT binding_site 495..498
FT /note= "Asn-X1-X2-Ser motif (X1-Ile, Leu or Tyr
FT and X2-Val, Ile or Ala)"
FT region 804..809
FT /note= "strongly charged C-terminal region"
FR F2709132-A.
PN PN
PD 24-FEB-1995.
PF 17-AUG-1993; Q10050.
PR 17-AUG-1993; FR-010050.
PA (SNFI ) ELF SANOFI.
PI Legoux R, Lelong P, Salome MLV;
DR WPI: 95B-100554/14.
DR N-PSDB; Q86268.
PT New DNA encoding enzyme for fragmenting N-acetyl-heparosan -
PT giving products useful as pharmaceutical starting material, also
PT related protein, vector, transformed cells, etc.
PS Claim 2; Page 38-40; 54pp; French.
CC An enzyme which degrades high mol.wt. N-acetylheparosan into
CC fragments of lower mol.wt. (more suitable for use as pharmaceutical
CC starting materials) can be obtained from E.coli (K5) SEBR 3282. The
CC DNA sequence (Q86268) coding for the enzyme was isolated in a
CC plasmid (p38.7) which has been deposited in E.coli RL1 strain as
CC CMCN I-1352. The deduced amino acid sequence (R71802) has homology
CC with an exo-poly-alpha-D-galacturonidase from Erwinia chrysanthemi.
SQ Sequence 820 AA;

Query Match 30.2%; Score 70; DB 13; Length 820;
Best Local Similarity 34.3%; Pred. No. 4.14e+01;
Matches 12; Conservative 9; Mismatches 13; Indels 1; Gaps 1;

Db 197 gvsikfagqngilndnk-d-aftkshsfsvfp 230
|: |: | | | | | | | | | | | | | |
QY 59 GLHLRVDPSTGALVDSKSYAFSTNDTTSAAFVS 93

RESULT 5
```

```

DR WPI: 98-542705/46.
DR N-PSDB: V63945.
PT New isolated mycobacteria polypeptides and nucleic acids - used for
PT developing products for the diagnosis of or vaccination against
PT mycobacterial infections, particularly tuberculosis
PS Claim 1; Page 223; 163pp; English.
CC The present sequence represents a Mycobacterium tuberculosis protein.
CC Products from the present invention, which describes protein fragments
CC and nucleic acid fragments derived from M. tuberculosis, can be used in
CC the detection of and prevention of mycobacterial infections. In
CC particular, the proteins and nucleic acids can be used for the diagnosis
CC of or vaccination against tuberculosis caused by M. tuberculosis,
CC M. africanum or M. bovis.
SQ Sequence 129 AA;

      Query Match      29 38; Score 68; DB 37; Length 129;
      Best Local Similarity 26.58; Pred. No. 6.11e-01;
      Matches 9; Conservative 12; Mismatches 12; Indels 1; Gaps 1;

Db 9 ltrllaasagilaa-afaaqaapgvddafia 41
   | | | : | | | : | | : : | | : :
QY 60 LHLVFPSTGALVDSKSYAFSTNDTTNAFVS 93

RESULT 7
ID R80530 standard; Protein; 1252 AA.
AC R80530;
DT 22-DEC-1995 (first entry)
DE B. sphaericus SLP.
KW Surface layer protein; SLP; fusion protein; vaccine; antigen;
KW surface expression; epitope.
OS Bacillus sphaericus.
FH Key Location/Qualifiers
FT 1-30
FT peptide /label= sig_peptide
FT WO9519371-A2.
PN 20-JUL-1995.
PD PD
PF 13-JAN-1995; E00147.
PF 14-JAN-1994; GB-000650.
PA (SOLV ) SOLVAY SA.
PI Deblaere RY, Desomer J, Dhaese P;
DR WPI: 95-263827/34.
DR N-PSDB: Q99430.
PT Host cell expressing surface layer protein fusion protein - used for
PT host presentation of antigens and vaccine prodrn.
PS Disclosure; Fig.6; 95pp; English.
CC A probe based on the N-terminal sequence of B. sphaericus P-1 (LMG
CC p-13955) surface layer protein was used to screen an HindIII-
CC generated library to isolate the slp gene. Promoter regions
CC of the gene are used in genetic constructs providing surface
CC expression of heterologous proteins in P-1 hosts.
SQ Sequence 1252 AA;

```

Db	47	tvelnlkeayqpalaelaemprpyvfvs	76
	:	:   :	:   :
Qv	57	SAGLHLR-VFDPSTGALVDSKSYAFST	85

```
Query Match      29.3% Score 68; DB 14; Length 1252;  
Best Local Similarity 31.3%; Pred.No. 6,lle-01;  
Matches 10; Conservative 12; Mismatches 8; Indels 2; Gaps 2;
```

---

```
Dbb       778 vkviatgiavntdkgdyar-takeat-aftta 807  
          :|:: :: |:: |||||::|::|::|:  
QY        62 LRVEDSTGALVDKSKYAFTSNDTTSAAFSV 93
```

RESULT	8
ID	W41513 standard; Protein; 371 AA.
AC	W41513;
DT	17-JUN-1998 (first entry)
DE	N. meningitidis alpha-2,3-sialyltransferase protein.
KW	Alpha-2,3-sialyltransferase; sialic acid; acceptor; lipid;
KW	biologically active oligosaccharide; sialyl-modified protein.
OS	Neisseria meningitidis.
PN	W09747749-AL.
PD	18-DEC-1997.
PF	10-JUN-1997; CA0390.

```

PR 10-JUN-1997: US-872485.
PR 10-JUN-1996: US-019520.
PA (CANADA) NAT RES COUNCIL CANADA.
PI Gilbert M, Jennings MP, Wakarchuk WW, Young NM;
DR WPI; 98-052313/05.
PR N-PSDB: V04125.
PT Nucleic acid sequence encoding Neisseria
PT alpha-2,3-sialyltransferase - useful to add sialic acid to acceptor
PT with terminal galactose residue for synthesis of biologically active
PT oligosaccharide 38; 50pp; English.
PS Claim 31; Page 38; 50pp; English.
CC This sequence represents an alpha-2,3-sialyltransferase which has been
CC isolated from Neisseria meningitidis. The protein can be used as a
CC reagent for adding a sialic acid residue to an acceptor having a
CC terminal galactose residue, e.g. in synthesis of biologically active
CC oligosaccharides or sialyl-modified proteins or lipids. The nucleic
CC acid sequence can be used for the recombinant production of
CC alpha-2,3-sialyltransferase.
SQ Sequence 371 AA;

Query Match 28.9%; Score 67; DB 29; Length 371;
Best Local Similarity 47.1%; Pred. No. 7.42e+01;
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 160 iktfdgtgnlqssy 176
QY 62 LRVDPSTGALVDSKSY 78

RESULT 9
ID W21704 standard; Protein; 248 AA.
AC W21704;
DT 26-SEP-1997 (first entry)
DE Luffin-A.
KW pro-Ribosome inactivating protein; proRIP; peptide linker; cancer;
KW inactivation; eukaryotic ribosome; alpha fragment; beta fragment;
KW inhibitor; protein synthesis; N-glycosidase; glycosidic bond; liver;
KW rat; ribosomal 28S RNA; cellular proliferation; HIV-infected T cell.
OS Luffa cylindrica.
FH Key Location/Qualifiers
FT region 138..148
FT /note= "Position of possible insertion of internal
FT peptide linker sequence"

US5635384-A.
PN 03-JUN-1997.
PF 11-JUN-1990; US-535636.
PR 11-JUN-1990; US-535636.
PR 09-DEC-1992; US-987927.
PR 26-JAN-1995; US-378761.
PA (DOMC) DOWELANCO.
PI Hey TD, Morgan AER, Walsh TA;
DR WPI; 97-309831/28.
PT Inactive precursor of maize ribosome-inactivating protein - also
PT chimeric ribosome-inactivating protein precursors containing
PT internal linker sequences
PS Claim 2; Column 117-118; 121pp; English.
CC The sequences given in W21698-710 represent Ribosome Inactivating
CC proteins (RIP's), which may be used in the construction of the
CC proRIP of the invention. The proRIP has a selectively removable,
CC internal peptide linker. The precursor sequence is incapable of
CC inactivating eukaryotic ribosomes, but can be converted by removal
CC of the linker into a protein having alpha and beta fragments and being
CC capable of inactivating eukaryotic ribosomes. RIPs are potent
CC inhibitors of eukaryotic protein synthesis. They possess a highly
CC specific N-glycosidase activity which cleaves the glycosidic bond of
CC adenine 4324 of rat liver ribosomal 28S RNA. RIP's selectively inhibit
CC cellular proliferation of cells, e.g. cancer cells and HIV-infected T
CC cells. The inactive proRIP proteins make it possible to provide protein
CC synthesis inhibitors with uses in practical and improved ways not before
CC possible. The RIP can be used to make cytotoxic conjugates.
SQ Sequence 248 AA;

Query Match 28.4%; Score 66; DB 23; Length 248;
Best Local Similarity 39.1%; Pred. No. 8.99e+01;
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Db 73 gylvnstsyffnesdaklasqyv 95
QY 70 GALVDSKSYAFSTNDTSAAFV 92

RESULT 11
ID R74180 standard; Protein; 248 AA.
AC R74180;
DT 01-JAN-1996 (first entry)
DE Type I Ribosome-inactivating protein luffin.
KW Ribosome inactivating protein; RIP; Type I; cytotoxin; immunotoxin.
OS Plant.
FH Key Location/Qualifiers
FT misc_difference 14
FT /label= invariant residue
FT /note= "in ricin A-chain and the Type I RIPs"
FT misc_difference 22
FT /label= see above
FT misc_difference 70
FT /label= see above
FT misc_difference 110
FT /label= see above
FT misc_difference 131
FT /label= see above
FT misc_difference 157..158
FT /label= see above
FT misc_difference 160
FT /label= see above
FT misc_difference 192
FT /label= see above

```

```
PN US5416202-A.
PD 16-MAY-1995.
PF 09-DEC-1992; 988430.
PR 04-NOV-1991; US-787567.
PR 09-DEC-1992; US-988430.
PA (XOMA ) XOMA CORP.
PI Bernhard SL, Better MD, Carroll SF, Lane JA, Lei SP;
DR WPI; 95-193480/25.
PT Polynucleotide(s) encoding gelonin analogues - having a cystein
PT residue for intermolecular bonding for the prodn. of immuno-toxins(s)
PS Disclosure; Figure 4; 66pp; English.
CC Analogues of type I RIP are defined as non-naturally occurring
CC polypeptides that share the ribosome-inactivating activity of the
CC natural protein but differ in AA sequence. Preferred analogues have
CC a Cys available for disulfide bonding located at a posn. it its AA
CC sequence from the posn. corresp. to posn. 251 in ricin A-chain RTA
CC to the carboxy terminus of the analogue. (R74176 is the sequence of
CC ricin A-chain RTA, which is a Type II RIP). The primary AA
CC sequence of the Type I RIPs gelonin, BRIP, mormordin II, luffin
CC [see Islam et al., Agricultural Biological Chem., 54(5) 1343-45 1991],
CC alapharichosanthin [fsee Chow et al., J. Biol. Chem., 265, 8670-74
CC 1990], momordin I [see Ho et al., BBA, 1088, 311-14 1991],
CC Mirabilis antiviral protein [see Habuka et al., J. Biol. Chem.,
CC 264(12) 6629-37 1989], pokeweed antiviral protein isolated from
CC seeds [see Kung et al., Agric. Biol. Chem., 54(12), 3301-18 1990]
CC and saporin [see Banatti et al., Eur. J. Biochem., 183, 465-70
CC 1989] are individually aligned with the primary sequence of the
CC ricin A-chain [see Halling et al., Nucleic Acids Res., 13,
CC 8019-8033 1985] respectively in Figures 1-9. The AAs invariant
CC among the ricin A-chain and the Type I RIPs are indicated in FT.
SQ Sequence 248 AA;

Query Match 28.4%; Score 66; DB 14; Length 248;
Best Local Similarity 39.1%; Pred. No. 8,99e+01;
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Db 73 gylvnstsyyfnesdaklasqyv 95
QY 70 GALVDSKSYAFSTSDNTTSAEFV 92

RESULT 12
ID R12468 standard; Protein; 248 AA.
AC R12468;
DT 29-AUG-1991 (first entry)
DE Luffa cylindrica bioactive protein.
KW Protein synthesis inhibition.
OS Luffa cylindrica.
PN J03109398-A.
PD 09-MAY-1991.
PR 22-SEP-1989; 247410.
PR 22-SEP-1989; JP-247410.
PA (SUMO ) SUMITOMO CHEM IND KK.
DR WPI; 91-181458/25.
PT New protein of Luffa cylindrica, extracted from seeds - is used for
PT inhibiting protein synthesis in viral proliferation.
PS Claim 1; Fig 1; 6pp; Japanese.
CC The protein is prepd. by extraction from seeds of L. cylindrica.
CC It has protein synthesis inhibiting activity and is useful for
CC inhibiting viral growth.
CC See also J0112999.
SQ Sequence 248 AA;

Query Match 28.4%; Score 66; DB 3; Length 248;
Best Local Similarity 39.1%; Pred. No. 8,99e+01;
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Db 73 gylvnstsyyfnesdaklasqyv 95
QY 70 GALVDSKSYAFSTSDNTTSAEFV 92

RESULT 13
ID R12468 standard; Protein; 248 AA.
AC R12468;
DT 29-AUG-1991 (first entry)
DE Luffa cylindrica bioactive protein.
KW Protein synthesis inhibition.
OS Luffa cylindrica.
PN J03109398-A.
PD 09-MAY-1991.
PR 22-SEP-1989; 247410.
PR 22-SEP-1989; JP-247410.
PA (SUMO ) SUMITOMO CHEM IND KK.
DR WPI; 91-181458/25.
PT New protein of Luffa cylindrica, extracted from seeds - is used for
PT inhibiting protein synthesis in viral proliferation.
PS Claim 1; Fig 1; 6pp; Japanese.
CC The protein is prepd. by extraction from seeds of L. cylindrica.
CC It has protein synthesis inhibiting activity and is useful for
CC inhibiting viral growth.
CC See also J0112999.
SQ Sequence 248 AA;

Query Match 28.4%; Score 66; DB 26; Length 248;
Best Local Similarity 39.1%; Pred. No. 8,99e+01;
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Db 73 gylvnstsyyfnesdaklasqyv 95
QY 70 GALVDSKSYAFSTSDNTTSAEFV 92

RESULT 14
ID W25141 standard; Protein; 248 AA.
AC W25141;
DT 02-DEC-1997 (first entry)
DE Luffin-A (a ribosome inhibitory protein) inactive precursor.
KW Maize; proKIP; ribosome inactivating protein; alpha; beta subunit;
KW Internal linker; Barley Translation Inhibitor; Trichosanthin;
KW Ricin A-chain; Abrin-A A-chain; Saporin; SLT-1; Luffin-A; MAP;
KW Ricinus communis agglutinin; Momordin; PAP-S; Luffin-B; Dianthin 30;
KW therapeutic toxin; tumour cell targeted; protein synthesis inhibitor;
KW post-translational modification; cancer; neoplasia; HIV; AIDS;
KW human immunodeficiency virus; acquired immune deficiency syndrome.
OS Synthetic.
PN US5646026-A.
PD 08-JUL-1997.
PR 11-JUN-1990; 535636.
PR 09-DEC-1992; US-987927.
PR 11-JUN-1990; US-535636.
PR 26-JAN-1995; US-378761.
PR 07-JUN-1995; US-485286.
PA (DOWC ) DOWELANCO.
PI Hey TD, Morgan AER, Walsh TA;
DR WPI; 97-362934/33.
PT DNA encoding pro-ribosome inactivating proteins - inactive
PT precursors of ribosome inactivating proteins; can be expressed in
PT eukaryotic cells without causing cell death
PS Claim 4; Column 117-120; 186pp; English.
CC W25141 shows a Luffin-A (ribosome inhibitory protein, RIP) protein
CC which was engineered to contain a selectively removable internal peptide
CC linker sequence separating the alpha and beta units of the RIP. When
CC separated the two units regain activity and are capable of inactivating
CC eukaryotic ribosomes and hence preventing protein production. Many
CC different RIPs may be produced with an internal linker including
```

CC maize RIP, Trichosanthin, Ricin A-chain, Abrin-A A-chain and  
 CC Saporin. The RIPs can be used in the construction of therapeutic  
 CC toxins targeted to specific cells such as tumour cells via the  
 CC attachment of a targeting polypeptide, e.g. a monoclonal antibody.  
 CC A further use is in HIV therapy (see US4869903). There is interest  
 CC in expressing RIP recombinantly in host eukaryotic cells, because of  
 CC the capacity to provide correct post-translational processing. However,  
 CC RIPs effectively inhibit protein synthesis in eukaryotic cells resulting  
 CC in cell death. Since the inactive RIP proteins are not cytotoxic to  
 CC eukaryotic cells, they can be recombinantly expressed in such cells and  
 CC then converted to active RIP proteins.  
 SQ Sequence 248 AA;

Query Match 28.4%; Score 66; DB 24; Length 248;

Best Local Similarity 39.1%; Pred. No. 8.99e+01;

Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Db 73 gylvnstsyffnesdaklasqv 95

QY 70 GALVDSKSYAFSTNDTTSAAFV 92

# RESULT 15

ID R53731 standard; Protein: 276 AA.

AC R53731;

DT 09-FEB-1995 (first entry)

DE Luffin (ribosome inactivating protein).

KW Luffin; ribosome inactivating protein; RIP; loofah; inhibition;

KW protein translation; protein synthesis; viral proliferation.

OS Luffa cylindrica.

FH Key location/Qualifiers

FT peptide 1..20

FT protein /label= signal\_peptide

FT FT 21..276

FT FT /label= luffin

FT misc\_difference 123

FT /note= "Leu corresponds to TCC codon"

PN J06141867-A.

PD 24-MAY-1994.

PF 04-NOV-1992; 295301.

PR 04-NOV-1992; JP-295301.

PA (TORA ) TORAY IND INC.

DR WPI; 94-205026/25.

DR N-PSDB; Q66093.

PT DNA encoding luffin - used for mass production of luffin, a

PT ribosome inactivating protein

PS Claim 2; Page 4-5; 5pp; Japanese.

CC The luffin coding sequence was isolated from a cDNA library prepared

CC from mRNA isolated from seeds of Luffa cylindrica. The cDNA will be

CC useful for recombinant production of luffin, a ribosome inactivating

CC protein. The protein inhibits protein synthesis and viral

CC replication.

SQ Sequence 276 AA;

Query Match 28.4%; Score 66; DB 10; Length 276;

Best Local Similarity 39.1%; Pred. No. 8.99e+01;

Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Db 92 gylvnstsyffnesdaklasqv 114

QY 70 GALVDSKSYAFSTNDTTSAAFV 92

Search completed: Wed Sep 1 16:26:18 1999  
 Job time : 38 secs.



```

REFERENCE S45656
#authors Skala, J.; van Dyck, L.; Purnelle, B.; Goffeau, A.
#journal Yeast (1994) 10(Suppl.A):S13-S24
#title The sequence of an 8.8 kb segment on the left arm of chromosome II from Saccharomyces cerevisiae reveals four new open reading frames including homologs of animal DNA polymerase alpha-primases and bacterial GTP cyclohydrolase II.

#accession S45656
##molecule_type DNA
##residues 1-345 ##label SKA
##cross-references EMBL:X74738; NID:g5111140; PID:g5111141
##experimental_source strain S288C
#REFERENCE S50971
#authors Saiz, J.; Santos, M.A.; Plaza, M.A.; Revuelta, J.L.
#submission Submitted to the EMBL Data Library, February 1993
#description Cloning and sequencing of the RIB1 gene from Saccharomyces cerevisiae.
#accession S50971
##molecule_type DNA
##residues 1-180,'N',182-345 ##label SAI
##cross-references EMBL:Z21617; NID:g642218; PID:g642219
#COMMENT This enzyme is involved in riboflavin (vitamin B2) biosynthesis, catalyzing the release of formate and inorganic pyrophosphate under formation of 2,5-diamino-6-ribosylamino-4(3H)-pyrimidine 5'-phosphate from GTP.
#GENETICS
#gene SGD:RIB1
##cross-references SGD:S0000129; MIPS:YBL033c
#map_position 2L
#CLASSIFICATION #superfamily yeast cyclohydrolase II; cyclohydrolase homology
#KEYWORDS GTP; hydrolase; riboflavin biosynthesis
#FEATURE
139-309
#domain cyclohydrolase homology #label CYCH
#length 345 #molecular-weight 38346 #checksum 7596
SUMMARY
Query Match 31.5%; Score 73; DB 2; Length 345;
Best Local Similarity 50.0%; Pred. No. 2.75e+00;
Matches 14; Conservative 4; Mismatches 9; Indels 1; Gaps 1;
Db 114 GLALE-FDDSTGELLASKATTWDAHNPT 140
||| ||| ||| : ||| : |||
QY 59 GLHLRVDPSTGALVDSKSYAFSTSNPT 86
||||| ||| ||| : ||| : |||

RESULT 3
ENTRY H69418 #type complete
TITLE hypothetical protein AF1353 - Archaeoglobus fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
ACCESSION H69418
REFERENCE H69250
#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickley, E.K.; Peterson, J.D.; Richardson, D.B.; Kerlavage, A.R.; Graham, D.E.; Kyriakides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Cocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs, T.; Artachon, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
#journal Nature (1997) 390:364-370
#title The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MUID:98049343
#accession H69418
##status preliminary; nucleic acid sequence not shown; translation not shown

```





```
##experimental_source strain S288C
REFERENCE S72107
#authors Eide, L.G.; Sander, C.; Prydz, H.
#journal Yeast (1996) 12:1085-1090
#title Sequencing and analysis of a 35.4 kb region on the left arm
of chromosome IV from Saccharomyces cerevisiae reveal 23
open reading frames.
#cross-references MUID:97051598
#accession S72126
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-889 #label EIW
##cross-references EMBL:X95966; NID:g1216215; PID:g225538; PID:g1216235
##note the nucleotide sequence was submitted to the EMBL Data
Library, February 1996
GENETICS
#map_position 4R
#note YDR027C
KEYWORDS transmembrane protein
FEATURE
767-783
#domain transmembrane #status predicted #label TMW
#length 889 #molecular-weight 101519 #checksum 1524
SUMMARY
Query Match 31.08; Score 72; DB 2; Length 889;
Best Local Similarity 31.08; Pred. No. 3.86e+00;
Matches 9; Conservative 8; Mismatches 12; Indels 0; Gaps 0;
Db 77 RSPDSRTLSVANSRSGFENETHSGSMDF 105
| | | | | : | | | | | : | | | | |
QY 63 RVDPSTGALVDSKSYAFSTNDTTSAAF 91

RESULT 10
ENTRY H36812
TITLE #type complete
ORGANISM hypothetical protein ORF64 - saimirine herpesvirus 1 (strain
11)
#formal_name saimirine herpesvirus 1
#note host Saimiri sciureus (common squirrel monkey)
DATE 16-Oct-1992 #sequence_revision 16-Oct-1992 #text_change
09-Sep-1997
ACCESSIONS H36812
REFERENCE H36806
#authors Albrecht, J.
#submission Submitted to the EMBL Data Library, January 1992
#description Primary structure of the herpesvirus saimiri genome.
#accession H36812
##molecule_type DNA
##residues 1-2469 #label ALB
##cross-references GB:X64346; NID:g60320; PID:g60385
REFERENCE A37309
#authors Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.;
Biesinger, B.; Newman, C.; Wittmann, S.; Craxton, M.A.;
Coleman, H.; Fleckenstein, B.; Honess, R.W.
#journal J. Virol. (1992) 66:5047-5058
#title Primary structure of the herpesvirus saimiri genome.
#cross-references MUID:92333688
#contents annotation; protein-coding frames
#note neither protein nor nucleotide sequence is given
GENETICS
#gene 64
#summary #length 2469 #molecular-weight 280165 #checksum 4051
Query Match 31.08; Score 72; DB 2; Length 2469;
Best Local Similarity 38.78; Pred. No. 3.86e+00;
Matches 12; Conservative 8; Mismatches 10; Indels 1; Gaps 1;
Db 161 IFDPHSCPLVPNSPAHVFSNVDATLEYLS 191
| | | | | : | | | | | : | | | | |
QY 64 VFDPSTGALV-DSKSYAFSTNDTTSAAFVS 93

RESULT 11
ENTRY D69200
#type complete
#experimental_source strain MTH751 - Methanobacterium
thermoautotrophicum (strain Delta H)
#formal_name Methanobacterium thermoautotrophicum
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
04-Sep-1998
ACCESSIONS A69200
REFERENCE A69200
#authors Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, P.;
Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.; Caruso,
A.; Bush, D.; Sifer, H.; Patwell, D.; Prabhakar, S.;
McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.;
Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
J.; Reeve, J.N.
#journal J. Bacteriol. (1997) 179:7135-7155
#title Complete genome sequence of Methanobacterium
thermoautotrophicum Delta H: functional analysis and
comparative genomics.
#cross-references MUID:98037514
#accession D69200
##status preliminary: nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-241 #label MTH
##cross-references GB:AE000854; GB:AE000666; NID:g2621839; PID:g2621840
##experimental_source strain Delta H
GENETICS
#gene MTH751
#start_codon GTG
CLASSIFICATION #superfamily hypothetical protein MJ1374
SUMMARY #length 241 #molecular-weight 26734 #checksum 570
Query Match 30.68; Score 71; DB 2; Length 241;
Best Local Similarity 38.98; Pred. No. 5.39e+00;
Matches 14; Conservative 8; Mismatches 11; Indels 3; Gaps 2;
Db 24 AGNIIHI-DPGCALV--RSYQFSDPRKLDAYMVS 56
| | | | | : | | | | | : | | | | |
QY 58 AGHLRVFDPSTGALVDSKSYAFSTNDTTSAAFVS 93

RESULT 12
ENTRY S69889
TITLE #type complete
ORGANISM hemagglutinin precursor - influenza A virus (isolate
sw/Ehime/1/80 [H1N2])
#formal_name influenza A virus
#variety H1N2 isolate sw/Ehime/1/80
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
01-Feb-1999
ACCESSIONS S69889
REFERENCE S69886
#authors Sugita, S.; Yoshioka, Y.; Itamura, S.; Kanegae, Y.; Oguchi,
K.; Gojobori, T.; Nerome, K.; Oya, A.
#journal J. Mol. Evol. (1991) 32:16-23
#title Molecular evolution of hemagglutinin genes of H1N1 swine and
human influenza A viruses.
#cross-references MUID:911866408
#accession S69889
##status nucleic acid sequence not shown
##molecule_type genomic RNA
##residues 1-566 #label SUG
##cross-references EMBL:X57494
GENETICS
#gene HA1; HA2
#classification #superfamily influenza virus hemagglutinin
#keywords glycoprotein; hemagglutinin; transmembrane protein
FEATURE
1-17
18-343
345-566
530-546
#domain signal sequence #status predicted #label SIG\
#product hemagglutinin HA1 #status predicted #label HA1\
#product hemagglutinin HA2 #status predicted #label HA2\
#domain transmembrane #status predicted #label TM1\
```

```

121-481,59-292,
72-84,153-488,
296-320
#disulfide_bonds #status predicted\
27,28,40,104,293,
304,498,557
#binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 566 #molecular-weight 63409 #checksum 2042
Query Match 30.6%; Score 71; DB 2; Length 566;
Best Local Similarity 37.9%; Pred. No. 5.39e+00;
Matches 11; Conservative 9; Mismatches 7; Indels 2; Gaps 2;

Db 259 FE-ATGNLVVPR-YAFAMNRDPGSGIIIS 285
I: || || :: ||| : | | : ||
QY 65 FDPSTGALVDSKSYAFSTNDTTSAAFVS 93

RESULT 13
ENTRY JT0345 #type complete
TITLE dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
ALTERNATE_NAMES (strain GS-5)
ORGANISM sucrose 6-glucosyltransferase
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
ACCESSIONS JT0345; C33135
REFERENCE JT0345
#authors Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
#journal Gene (1988) 69:101-109
#title Sequence analysis of the gtfC gene from Streptococcus mutans
GS-5.
#cross-references MUID:89137980
#accession JT0345
#molecule_type DNA
#residues 1-1375 #label UED
#experimental_source GS-5
REFERENCE A33135
#authors Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
#journal J. Bacteriol. (1987) 169:4263-4270
#title Sequence analysis of the gtfB gene from Streptococcus mutans.
#cross-references MUID:87308013
#accession C33135
#status preliminary
#molecule_type DNA
#residues 1-349 #label SHI
#cross-references GB:M17361
GENETICS #gene gtfC
FUNCTION #description catalyzes the synthesis of both water-soluble and
water-insoluble glucans from glucose.
CLASSIFICATION #superfamily cpl repeat homology
KEYWORDS duplication; glycosyltransferase; hexosyltransferase;
transferase
FEATURE
1-34 #domain signal sequence #status predicted #label SIG\
35-1375 #product glucosyltransferase #status predicted #label
MAR\
1126-1145 #domain cpl repeat homology #label CP1\
1253-1272 #domain cpl repeat homology #label CP2\
1318-1337 #domain cpl repeat homology #label CP3
SUMMARY #length 1375 #molecular-weight 153021 #checksum 7015
Query Match 30.6%; Score 71; DB 2; Length 1375;
Best Local Similarity 41.4%; Pred. No. 5.39e+00;
Matches 12; Conservative 8; Mismatches 7; Indels 2; Gaps 2;

Db 1093 SSVTGLVFDGKGYVYSTSGNQAKAFIS 1121
I: || || ||| : ||| : ||| : |||
QY 67 PSTGALV-DSKSYAF-STSDNTTSAAFVS 93

RESULT 14

```

```

ENTRY S25345 #type complete
TITLE probable membrane protein YCR089w - yeast (Saccharomyces
cerevisiae)
ALTERNATE_NAMES hypothetical protein YCR1102
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
06-Feb-1998
ACCESSIONS S25345; S19504
REFERENCE S25345
#authors Wilson, C.; Grisanti, P.; Frontali, L.
#journal Yeast (1992) 8:569-575
#title The complete sequence of a 6146 bp fragment of Saccharomyces
cerevisiae chromosome III contains two new open reading
frames.
#cross-references MUID:92397594
#accession S25345
#molecule_type DNA
#residues 1-1609 #label WIL
#cross-references GB:X59720; EMBL:S43845; NID:g1907116; PID:e264634;
PID:g1907227
REFERENCE S19504
#authors Frontali, L.; Grisanti, P.
#submission submitted to the Protein Sequence Database, March 1992
#accession S19504
#molecule_type DNA
#residues 1-1609 #label PRO
#cross-references EMBL:X59720; NID:g1907116; PID:e264634; PID:g1907227;
MIPS:YCR089w
GENETICS #gene SGD:FIG2
#cross-references SGD:S0000685; MIPS:YCR089w
#map_position 3R
KEYWORDS transmembrane protein
FEATURE
4-20 #domain transmembrane #status predicted #label TM1\
1592-1609 #domain transmembrane #status predicted #label TM2
SUMMARY #length 1609 #molecular-weight 166047 #checksum 8346
Query Match 30.6%; Score 71; DB 2; Length 1609;
Best Local Similarity 34.2%; Pred. No. 5.39e+00;
Matches 13; Conservative 12; Mismatches 12; Indels 1; Gaps 1;

Db 145 SSSFELPVPAPSSSLPSSSTSLTFTSVNPSSQMTSFS 182
I: :: | | | :: | | | :: | | | :: |
QY 57 SAGLHLRVDPSTGALVDSKSYAFSTNDTTS-AAFVS 93

RESULT 15
ENTRY S30708 #type complete
TITLE hypothetical 15.8K protein (corA-rarD intergenic region) -
Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 02-Dec-1993 #sequence_revision 01-Sep-1995 #text_change
14-Nov-1997
ACCESSIONS S30708; S30745; C65186
REFERENCE S30660
#authors Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.
#journal Science (1992) 257:771-778
#title Analysis of the Escherichia coli genome: DNA sequence of the
region from 84.5 to 86.5 minutes.
#cross-references MUID:92358234
#accession S30708
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-138 #label DAN
#cross-references EMBL:M87049; NID:g836656; PID:g148217
#note the nucleotide sequence was submitted to the EMBL Data
Library, November 1992
REFERENCE S30742
#authors Ohmori, H.
#submission submitted to the EMBL Data Library, January 1993
#description Physical map of the corA region of the E.coli chromosome.

```

```

#accession      S30745
#status         preliminary
#molecule_type DNA
#residues       1-138 ##label OHM
##cross-references EMBL:L02122; NID:g145575; PID:g145579
REFERENCE
#authors        Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
                Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
                Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Shao,
                Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.;
                Y.
#journal         Science (1997) 277:1453-1462
#title          The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession      C65186
#status         preliminary; nucleic acid sequence not shown;
                translation not shown
#molecule_type DNA
#residues       1-138 ##label BLAT
##cross-references GB:AE000457; GB:U00096; NID:g2367294; PID:g1790251;
                UWGP:b3818
##experimental_source strain K-12, substrain MG1655
GENETICS
#gene           yigG
#map_position   86 min
SUMMARY         #length 138 #molecular-weight 15794 #checksum 1118
Query Match     30.2%; Score 70; DB 2; Length 138;
Best Local Similarity 35.0%; Pred. No. 7.50e+00;
Matches         7; Conservative 6; Mismatches 7; Indels 0; Gaps 0;
Db              14 LRIFPTSGKISRRRIYFS 33
                ||| |::: | |
QY              62 LRVFDPSTGALVDSKSYAFS 81

```

Search completed: Wed Sep 1 16:25:22 1999  
Job time : 22 secs.

\*\*\*\*\*  
WATERMAN  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Sep 1 16:23:45 1999; MasPar time 3.53 Seconds  
Tabular output not generated. 296.202 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (57-93) from PCTUS9913024.ppt (9 of 12)  
Perfect Score: 232  
Sequence: 1 SAGLHLRVDPSTGALVDSKSYAFSTNDTSAFVS 37

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 33.585; Variance 52.447; scale 0.640

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	74	31.9	1222	1	YMH5_CAEEL	7.18e-01
2	73	31.5	345	1	GCH2_YEAST	1.04e+00
3	72	31.0	300	1	NTRY_AZOBR	1.50e+00
4	72	31.0	2469	1	TEGU_HSVSA	1.50e+00
5	71	30.6	1375	1	GTFC_STRMU	2.16e+00
6	71	30.6	1609	1	YIGG_ECOLI	2.16e+00
7	70	30.2	138	1	YIGG_ECOLI	3.10e+00
8	70	30.2	932	1	YMB4_CAEEL	3.10e+00
9	70	30.2	1545	1	IGB3_HAEIN	3.10e+00
10	70	30.2	1802	1	HKR1_YEAST	3.10e+00
11	69	29.7	566	1	FOBI_YEAST	4.42e+00
12	68	29.3	107	1	PRVB_ESOLU	6.29e+00
13	68	29.3	324	1	Y700_METJA	6.29e+00
14	67	28.9	153	1	HS11_SOYBN	8.91e+00
15	67	28.9	256	1	Y023_MYCTU	8.91e+00
16	67	28.9	533	1	VL2_HPV09	8.91e+00
17	67	28.9	639	1	CALC_RABIT	8.91e+00
18	67	28.9	1091	1	SYT_TREPA	8.91e+00
19	67	28.9	2431	1	POLN_SFV	8.91e+00
20	67	28.9	3063	1	COLLAGEN ALPHA 1(XII)	8.91e+00
21	67	28.9	3067	1	CALC_HUMAN	8.91e+00
22	66	28.4	108	1	PRVB_MERBI	1.26e+01
23	66	28.4	163	1	HMCS_DICDI	1.26e+01

24	66	28.4	277	1	RIPA_LUFYC	RIBOSOME-INACTIVATING	1.26e+01
25	66	28.4	539	1	TCPD_MOUSE	T-COMPLEX PROTEIN 1, D	1.26e+01
26	66	28.4	539	1	TCPD_HUMAN	T-COMPLEX PROTEIN 1, D	1.26e+01
27	66	28.4	577	1	SYN_HELPY	PROLYL-TRNA SYNTHETASE	1.26e+01
28	66	28.4	616	1	MUTA_STRCM	METHYLMALONYL-COA MUTA	1.26e+01
29	66	28.4	1430	1	GTFC_STRMU	GLUCOSYLTRANSFERASE-S	1.26e+01
30	65	28.0	108	1	PRVB_MERMR	PARVALBUMIN BETA.	1.76e+01
31	65	28.0	173	1	ELHA_APLCA	ATRIAL GLAND PEPTIDE A	1.76e+01
32	65	28.0	234	1	YODR_MYCTU	HYPOTHETICAL 25.4 KD P	1.76e+01
33	65	28.0	300	1	MOV_P_AMYA	MOVEMENT PROTEIN (CELL	1.76e+01
34	65	28.0	300	1	MOV_P_AMYE	MOVEMENT PROTEIN (CELL	1.76e+01
35	65	28.0	345	1	PURA_METJA	ADENYLOSUCCINATE SYNTH	1.76e+01
36	65	28.0	429	1	PURA_THIFE	ADENYLOSUCCINATE SYNTH	1.76e+01
37	65	28.0	480	1	TRA3_STAAT	TRANSPORIN FOR TRANS	1.76e+01
38	65	28.0	536	1	TCPD_FUGRU	T-COMPLEX PROTEIN 1, D	1.76e+01
39	65	28.0	554	1	VL2_HPV41	MINOR CAPSID PROTEIN L	1.76e+01
40	65	28.0	566	1	HEMA_TAZIN	HEMAGGLUTININ PRECURSO	1.76e+01
41	65	28.0	2710	1	TOXA_CLODI	COLD SHOCK-LIKE PROTEI	2.47e+01
42	64	27.6	68	1	CSPE_ECOLI	ASPARAGINE SYNTHETASE	2.47e+01
43	64	27.6	585	1	ASNI_LOTJA	HYPOTHETICAL 251.0 KD	2.47e+01
44	64	27.6	2167	1	YCS2_YEAST	DYNEIN BETA CHAIN, CIL	2.47e+01
45	64	27.6	4466	1	DYHC_TRIGR	DYNEIN BETA CHAIN, CIL	2.47e+01

ALIGNMENTS

RESULT 1  
ID YMH5\_CAEEL STANDARD; PRT: 1222 AA.  
AC P34472;  
DT 01-FEB-1994 (REL. 28, CREATED)  
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 136.3 KD PROTEIN F58A4.5 IN CHROMOSOME III.  
GN F58A4.5  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA;  
RN RHABDITIDA; RHABDITIDA; RHABDITIDA; PELODERINAE; CAENORHABDITIS.  
RC SEQUENCE FROM N.A.  
RX STRAIN-BRISTOL N2;  
RX MEDLINE; 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,  
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,  
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,  
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,  
RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,  
RA WATSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,  
RA WOLDMAN P.;  
RT \*2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RL elegans.;  
CC -1- SIMILARITY: TO REVERSE TRANSCRIPTASE.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; 222179; E1347484;  
DR PIR; S40977; S40977.  
DR WORMPEP; F58A4.5; CE00221.  
DR PFAM; PF00078; Ivt: 1.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 1222 AA; 136273 MW; 2F6802B7 CRC32;

Query Match 31.9%; Score 74; DB 1; Length 1222;

[2]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE; 92230228.
RT
RA NICHOLAS J., CAMERON K.R., COLEMAN H., NEWMAN C., HONESS R.W.;
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of
RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic

RT organization between HVS and Epstein-Barr virus.";  
RL Virology 188:296-310(1992).  
CC -!- FUNCTION: TEGUMENT PROTEIN.  
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,  
CC EHV-1 24, EBV BFLF1, HVS-1 64, VZV 22, AND HCMV UL48.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X64346; G60385; -  
CC EMBL; M86409; G330993; -  
CC PIR; H36812; H36812.  
CC SEQUENCE 2469 AA; 280165 MW; 1B72F5F9 CRC32;  
Query Match 31.0%; Score 72; DB 1; Length 2469;  
Best Local Similarity 38.7%; Pred. No. 1.50e+00;  
Matches 12; Conservative 8; Mismatches 10; Indels 1; Gaps 1;  
Db 161 IFDPHSCPVPNSPAHVSTSNVNDIAEVL 191  
QY 64 VEDPSIGALV-DSKSYAFSTSNDDTSAAFVS 93  
RESULT 5  
ID GTFC\_STRMU STANDARD; PRT: 1375 AA.  
AC P13470; P05427;  
DT 01-NOV-1988 (REL. 09, CREATED)  
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)  
DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).  
GN GTFC.  
OS STREPTOCOCCUS MUTANS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
OC STREPTOCOCCUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GS-5;  
RA MEDLINE; 89137980.  
RA UEDA S., SHIROZA T., KURAMITSU H.K.;  
RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5";  
RL GENE 69:101-109(1988).  
RN [2]  
RP SEQUENCE OF 1-349 FROM N.A.  
RC STRAIN-GS-5;  
RA MEDLINE; 87308013.  
RA SHIROZA T., UEDA S., KURAMITSU H.K.;  
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";  
RL J. BACTERIOL. 169:4263-4270(1987).  
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE  
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) -  
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).  
CC -!- SUBCELLULAR LOCATION: SECRETED.  
CC -!- DISEASE: DENTAL CARIES.  
CC -!- GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED  
CC GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE  
CC GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF  
CC GLUCANS.  
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
CC BINDING PROTEIN FROM S. MUTANS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial  
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M22054; G153643; -  
EMBL; M17361; G153641; -  
PIR; J00345; J00345.  
PIR; C33135; C33135.  
PFAM; PF00128; alpha-amylase; 1.  
KW TRANSFERASE; GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.  
FT SIGNAL 1 34  
FT CHAIN 35 1375 GLUCOSYLTRANSFERASE-SI.  
FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).  
FT DOMAIN 1126 1375 GLUCAN-BINDING (APPROXIMATE).  
FT REPEAT 1126 1375 2, 4 A, 1 C AND 1 AC REPEATS.  
FT REPEAT 1126 1159 A REPEAT.  
FT REPEAT 1159 1200 A REPEAT.  
FT REPEAT 1227 1238 C REPEAT.  
FT REPEAT 1253 1303 AC REPEAT.  
FT REPEAT 1318 1330 A REPEAT (INCOMPLETE).  
SQ SEQUENCE 1375 AA; 153022 MW; 3EA3727E CRC32;  
Query Match 30.6%; Score 71; DB 1; Length 1375;  
Best Local Similarity 41.4%; Pred. No. 2.16e+00;  
Matches 12; Conservative 8; Mismatches 7; Indels 2; Gaps 2;  
Db 1093 SSVTGLVDFGKGYVYVSTSGNOAKNAFIS 1121  
QY 67 PSTGALV-DSKSYAF-STSNDDTSAAFVS 93  
RESULT 6  
ID FIG2\_YEAST STANDARD; PRT: 1609 AA.  
AC P25653.  
DT 01-MAY-1992 (REL. 22, CREATED)  
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE FACTOR INDUCED GENE 2.  
GN FIG2 OR YCR089W OR YCR89W OR YCR1102.  
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;  
OC SACCHAROMYCETACEAE; SACCHAROMYCES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 92397594.  
RA WILSON C., GRISANTI P., FRONTALI L.;  
RT "The complete sequence of a 6146 bp fragment of Saccharomyces  
cerevisiae chromosome III contains two new open reading frames.";  
RL YEAST 8:569-575(1992).  
CC -!- FUNCTION: REQUIRED FOR EFFICIENT MATING.  
CC -!- INDUCTION: BY MATING PHEROMONES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X59720; E264634; -  
CC PIR; S19504; S19504.  
CC PIR; S25345; S25345.  
CC SGD; L0003312; FIG2.  
SQ SEQUENCE 1609 AA; 166049 MW; DE974CE8 CRC32;  
Query Match 30.6%; Score 71; DB 1; Length 1609;  
Best Local Similarity 34.2%; Pred. No. 2.16e+00;  
Matches 13; Conservative 12; Mismatches 12; Indels 1; Gaps 1;  
Db 145 SSSFELPVTAPSSSLPSTSLTFTSVNPSQSWTSFNS 182  
QY 57 SAGLHLRVDFPSTGALVDSKSYAFSTSNDDTTS-AAFVS 93

```

RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
RA SIMS M., SMALLON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOHLDMAN P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.*;
RL NATURE 368:32-38(1994).
CC -|- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -|- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: Z19155; G6715;
CC PIR: S28275; S28275;
CC WORPEP: F34G8.4; CE05942.
CC PROSITE: P500518; ZINC_FINGER_C3HC4; 1.
CC PFAM: PF00097; zf-C3HC4; 1.
CC HSP: P29590; 1BOR.
CC KW HYPOTHETICAL PROTEIN; ZINC-FINGER; DNA-BINDING; NUCLEAR PROTEIN.
CC ZN_FING 43 83 C3HC4-TYPE.
CC FT SEQUENCE 932 AA; 104438 MW; 45C877B4 CRC32;
CC
CC Query Match 30.2%; Score 70; DB 1; Length 932;
CC Best Local Similarity 34.4%; Pred. No. 3.10e+00;
CC Matches 11; Conservative 9; Mismatches 10; Indels 2; Gaps 2;
CC
Db 284 TEGEYLRNFQDADDELFRK-LRFSAGDDAA 314
QY 57 SAGLH-LRVFDPSTGALVDKSYAFSTNDTT 87

RESULT 9
ID IGA3_HAEIN STANDARD; PRT; 1545 AA.
AC P45385;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE IMMUNOGLOBULIN A1 PROTEASE PRECURSOR (EC 3.4.21.72) (IGAL PROTEASE).
GN IGA.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HK393 / NCTC 8467 / SEROTYPE B;
RX MEDLINE: 92234949.
RA POULSEN K., REINHOLDT J., KILIAN M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases.*;
RL J. BACTERIOL. 174:2913-2921(1992).
CC -|- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -|- CATALYTIC ACTIVITY: CLEAVAGE OF IMMUNOGLOBULIN A MOLECULES AT
CC CERTAIN PRO-1-XAA BONDS IN THE HINGE REGION. NO SMALL MOLECULE
CC SUBSTRATES ARE KNOWN.
CC -|- SUBCELLULAR LOCATION: SECRETED.
CC -|- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
RA SIMS M., SMALLON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOHLDMAN P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.*;
RL NATURE 368:32-38(1994).
CC -|- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -|- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: Z19155; G6715;
CC PIR: S28275; S28275;
CC WORPEP: F34G8.4; CE05942.
CC PROSITE: P500518; ZINC_FINGER_C3HC4; 1.
CC PFAM: PF00097; zf-C3HC4; 1.
CC HSP: P29590; 1BOR.
CC KW HYPOTHETICAL PROTEIN; ZINC-FINGER; DNA-BINDING; NUCLEAR PROTEIN.
CC ZN_FING 43 83 C3HC4-TYPE.
CC FT SEQUENCE 932 AA; 104438 MW; 45C877B4 CRC32;
CC
CC Query Match 30.2%; Score 70; DB 1; Length 932;
CC Best Local Similarity 34.4%; Pred. No. 3.10e+00;
CC Matches 11; Conservative 9; Mismatches 10; Indels 2; Gaps 2;
CC
Db 284 TEGEYLRNFQDADDELFRK-LRFSAGDDAA 314
QY 57 SAGLH-LRVFDPSTGALVDKSYAFSTNDTT 87

RESULT 9
ID IGA3_HAEIN STANDARD; PRT; 1545 AA.
AC P45385;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE IMMUNOGLOBULIN A1 PROTEASE PRECURSOR (EC 3.4.21.72) (IGAL PROTEASE).
GN IGA.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HK393 / NCTC 8467 / SEROTYPE B;
RX MEDLINE: 92234949.
RA POULSEN K., REINHOLDT J., KILIAN M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases.*;
RL J. BACTERIOL. 174:2913-2921(1992).
CC -|- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -|- CATALYTIC ACTIVITY: CLEAVAGE OF IMMUNOGLOBULIN A MOLECULES AT
CC CERTAIN PRO-1-XAA BONDS IN THE HINGE REGION. NO SMALL MOLECULE
CC SUBSTRATES ARE KNOWN.
CC -|- SUBCELLULAR LOCATION: SECRETED.
CC -|- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

RA DANIELS D.L., PLUNKETT G. III, BURLAND V.D., BLATTNER F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region
RT from 84.5 to 86.5 minutes.*;
RL SCIENCE 257:771-778(1992).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: L02122; G145579;
CC EMBL: M87049; G148217;
CC EMBL: AE000457; G1790251;
CC PIR: S30745; S30745.
CC PIR: S30708; S30708.
CC ECGENE: EG11465; YIGG.
CC KW HYPOTHETICAL PROTEIN.
CC SEQUENCE 138 AA; 15794 MW; C92E2E54 CRC32;
CC
CC Query Match 30.2%; Score 70; DB 1; Length 138;
CC Best Local Similarity 35.0%; Pred. No. 3.10e+00;
CC Matches 7; Conservative 6; Mismatches 7; Indels 0; Gaps 0;
CC
Db 14 LRIFITSNGKISRRIYIS 33
QY 62 LRVFDPSTGALVDKSYAFS 81

RESULT 8
ID YMB4_CAEEL STANDARD; PRT; 932 AA.
AC Q03601;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 104.4 KD PROTEIN F54G8.4 IN CHROMOSOME III.
GN F54G8.4.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTERA; RHABDITIDA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE OF 1-574 FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE: 94150718.
RA WILSON R., AINSKOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,

```

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M87490; G148909; - -  
 KW HYDROLASE; SERINE PROTEASE; TRANSMEMBRANE; ZYMOGEN; SIGNAL;  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 1012 IMMUNOGLOBULIN A1 PROTEASE.  
 FT PROPEP 1013 1545 HELPER PEPTIDE (POTENTIAL).  
 FT ACT SITE 292 292 PROBABLE.  
 SQ SEQUENCE 1345 AA; 170627 MW; 93EE4AD6 CRC32;

Query Match 30.2%; Score 70; DB 1; Length 1545;  
 Best Local Similarity 26.7%; Pred. No. 3.10e+00;  
 Matches 8; Conservative 14; Mismatches 6; Indels 2; Gaps 2;

Db 336 KIYEQISAGSLIGSKTIDYSSNSNGKTSIT 365

QY 63 RVFDP-STGALVDSKS-YAFSTSDNTTSA 90

RESULT 10  
 ID HKR1\_YEAST STANDARD; PRT: 1802 AA.  
 AC P41809;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE HANSENULA MRAKII KILLER TOXIN-RESISTANT PROTEIN 1 PRECURSOR.  
 GN HKR1 OR YDR420W.  
 OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 CC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;  
 CC SACCCHAROMYCETACEAE; SACCCHAROMYCES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-YNN 295;  
 RX MEDLINE; 94156857.  
 RA KASAHARA S., YAMADA H., MIO T., SHIRATORI Y., MIYAMOTO C.,  
 RA YABE T., NAKAJIMA T., ICHISHIMA E., FURUICHI Y.;  
 RT "Cloning of the Saccharomyces cerevisiae gene whose overexpression  
 RT overcomes the effects of HM-1 killer toxin, which inhibits  
 RT beta-glucan synthesis.";  
 RT J. BACTERIOL. 176:1488-1499(1994).  
 CC -!- FUNCTION: COULD REGULATE BETA-GLUCAN SYNTHESIS. OVEREXPRESSION  
 CC PROVIDES RESISTANCE TO HM-1 KILLER TOXIN.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE).  
 CC -!- PTM: COULD BE O-GLYCOSYLATED IN SERINE/THREONINE RICH DOMAIN.  
 CC -!- SIMILARITY: SOME, TO YEAST MSB2.  
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; S69101; G545660; - -  
 DR SGD; L0000788; HKR1  
 KW GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 1802 HANSENULA MRAKII KILLER TOXIN-RESISTANT  
 FT PROTEIN 1.  
 FT TRANSMEM 1486 1506 POTENTIAL.  
 FT DOMAIN 23 1478 SER/THR-RICH.  
 FT DOMAIN 453 788 12 X 28 AA TANDEM REPEATS OF S-[AV]-[P]-  
 FT Y-A-V-S-S-T-Y-T-S-S-P-A-P-A-I-S-S-T-  
 FT Y-T-S-S-P.  
 FT REPEAT 453 480 1 (APPROXIMATE).  
 FT REPEAT 481 508 2.  
 FT REPEAT 509 536 3.  
 FT REPEAT 537 564 4.

FT REPEAT 565 592 5.  
 FT REPEAT 593 620 6.  
 FT REPEAT 621 648 7.  
 FT REPEAT 649 676 8.  
 FT REPEAT 677 704 9.  
 FT REPEAT 705 732 10.  
 FT REPEAT 733 760 11.  
 FT REPEAT 761 788 12.  
 FT CARBOHYD 24 24 POTENTIAL.  
 FT CARBOHYD 1252 1252 POTENTIAL.  
 FT CARBOHYD 1293 1293 POTENTIAL.  
 FT CARBOHYD 1342 1342 POTENTIAL.  
 FT CARBOHYD 1400 1400 POTENTIAL.  
 SQ SEQUENCE 1802 AA; 188890 MW; 84B6C85F CRC32;

Query Match 30.2%; Score 70; DB 1; Length 1802;  
 Best Local Similarity 31.0%; Pred. No. 3.10e+00;  
 Matches 9; Conservative 12; Mismatches 7; Indels 1; Gaps 1;

Db 926 LSVNPSASNLVET-SLIISSTQASITSP 953

QY 62 LRVFDPSTGALVDSKSYAFSTSDNTTSA 90

RESULT 11  
 ID F0B1\_YEAST STANDARD; PRT: 566 AA.  
 AC O13329; 004587;  
 DT 15-JUL-1998 (REL. 36, CREATED)  
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
 DE DNA REPLICATION FORK BLOCKING PROTEIN F0B1.  
 GN F0B1 OR YDR110W OR YDR727.06.  
 OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 CC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;  
 CC SACCCHAROMYCETACEAE; SACCCHAROMYCES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C / AB972;  
 RA MURPHY L., SHORE L., HARRIS D., BARRELL B., RAJANDREAM M.A.,  
 RA WALSH S.V.;  
 RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.  
 CC -!- FUNCTION: ESSENTIAL FOR BOTH DNA REPLICATION FORK BLOCKING AND  
 CC RECOMBINATIONAL HOTSPOT ACTIVITIES.  
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF013245; G2331297; - -  
 DR EMBL; Z48758; G747885; - -  
 DR SGD; L0003959; F0B1.  
 FT CONFLICT 103 103 A -> T (IN REF. 2).  
 SQ SEQUENCE 566 AA; 65288 MW; 49D933EB CRC32;

Query Match 29.7%; Score 69; DB 1; Length 566;  
 Best Local Similarity 33.3%; Pred. No. 4.42e+00;  
 Matches 7; Conservative 8; Mismatches 5; Indels 1; Gaps 1;

Db 223 IHIEVFEPFNGEATGK-YSY 242

QY 60 LHLRVFDPSTGALVDSKSYAF 80









OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
OC STREPTOCOCCUS.  
[1]  
RN RN  
RP SEQUENCE FROM N.A.  
RC STRAIN-VI477;  
RA JAFFE R.I.;  
EL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR ENBL; AF049609; G2935546; -;  
FW TRANSFERASE.  
KT NON\_TER 1449 1449  
SQ SEQUENCE 1449 AA; 159895 MW; 2AE5FB6A CRC32;

Query Match            31.9%; Score 74; DB 2; Length 1449;  
Best Local Similarity 38.5%; Pred.No.1.77e+00;  
Matches 10; Conservative 5; Mismatches 11; Indels 0; Gaps 0

Dd 238 FNAETGELSNOKEYRFDKNGGTSSA 263  
      |::||::||::||::||::||:  
Qy 65 FDPSTGALVDSKSYAFSTSDNTTSAA 90

RESULT 5  
ID Q55264 PRELIMINARY; PRT: 1449 AA.  
AC Q55264;  
DT 01-NOV-1996 (TREMBREL\_01, CREATED)  
DT 01-NOV-1996 (TREMBREL\_01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBREL\_08, LAST ANNOTATION UPDATE)  
DE GLUCOSYLTRANSFERASE PRECURSOR.  
GN GFLL.  
OS STREPTOCOCCUS SALIVARIUS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
OC STREPTOCOCCUS.  
[1]  
RN RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95122197.  
RA SIMPSON C.L., GIFFARD P.M., JACQUES N.A.;  
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes  
coding for primer-independent glucosyltransferases.";  
RL INFECT. IMMUN. 63:609-621(1995).  
DR ENBL; L35495; G662379; -;  
DR PFAM; PF00128; alpha-amylase; 1.  
KW SIGNAL; TRANSFERASE.  
FT SIGNAL 1 35 POTENTIAL.  
FT CHAIN 36 1449 GLUCOSYLTRANSFERASE.  
SQ SEQUENCE 1449 AA; 159984 MW; E9A4BA87 CRC32;

Query Match            31.9%; Score 74; DB 2; Length 1449;  
Best Local Similarity 38.5%; Pred.No.1.77e+00;  
Matches 10; Conservative 5; Mismatches 11; Indels 0; Gaps 0

Dd 238 FNAETGELSNOKEYRFDKNGGTSSA 263  
      |::||::||::||::||::||:  
Qy 65 FDPSTGALVDSKSYAFSTSDNTTSAA 90

RESULT 6  
ID O28918 PRELIMINARY; PRT: 353 AA.  
AC O28918;  
DT 01-JAN-1998 (TREMBREL\_05, CREATED)  
DT 01-JAN-1998 (TREMBREL\_05, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBREL\_07, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 39.3 KD PROTEIN.  
GN AF1353.  
OS ARCHAEOLOBUS FULIGIDUS.  
OC ARCHAEA; EURYARCHAEOTA; ARCHAEOGLOBALES; ARCHAEOGLOBACEAE;  
OC ARCHAEOGLBUS.  
[1]  
RN RN  
RP SEQUENCE FROM N.A.  
RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE; 98049343.  
RA KLENN H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,  
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,  
RA RICHARDSON D.I., FERVAANCE A.B., GRAHAMMER,  
D.

[illegible]

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 ORF:PZF889.  
 DE SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OS EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMICETES; SACCHAROMYCETALES;  
 OC SACCHAROMYCETACEAE; SACCHAROMYCEDES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA EIDE L.G., SANDER C., PRYZD H.;  
 RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AB972;  
 RA BOWMAN S.;  
 RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AB972;  
 RA BARRELL B., RAJANDREAM M.A.;  
 RL SUBMITTED (JAN-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA PRYZD H., EIDE L.G.;  
 RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA MIPS;  
 RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.  
 DR EMBL; X95966; E225538; -  
 DR DRL; 247814; G642299; -  
 KR EMBL; 274323; E253298; -  
 KW HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 889 AA; 101519 MW; 3C15605E CRC32;

Query Match      31.0%; Score 72; DB 3; Length 889;  
 Best Local Similarity 31.0%; Pred. No. 3.58e+00;  
 Matches    9; Conservative    8; Mismatches 12; Indels    0; Ga

Db         77 RSEDFRTLSVANRSFGFTENGTHSGSDWF 105  
          |::|::|::|::|:  
 Qy         63 RVDPSTGALVDKSYAFSTDNTTSAAF 91  
          |||||::|::|:

RESULT          13  
 ID OZ6845 PRELIMINARY; PRT: 241 AA.  
 AC OZ6845;  
 DC 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
 GN CONSERVED PROTEIN.  
 GE MTH751.  
 OS METHANOBACTERIUM THERMOAUTOTROPHICUM.  
 OC ARCHAEA; EURYARCHAOTA; METHANOBAACTERIALES; METHANOBAACTERIACEAE;  
 CC METHANOBACTERIUM.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DELTA H;  
 RX MEDLINE; 98037514.  
 RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOISJS  
 RA ALDRIDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K,  
 RA HARRISON D., HOANG L., KEAGLE P., LUMB W., POTIER B., OTU D.,  
 RA SPADAFORA R., VICARE R., WANG Y., WIEZBOWSKI J., GIBSON R.,  
 RA JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRASHAKAR S.,  
 RA MCDOUGALL S., SHIMER G., GOVAL A., PIETROVSKI S., CHURCH G.M.,  
 RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVJE J.N.;  
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
 deltaH: functional analysis and comparative genomics."  
 RL J. BACTERIOLOG. 179:7135-7155(1997).  
 DR EMBL; AE000854; G2621840; -  
 SQ SEQUENCE 241 AA; 26734 MW; E6584334 CRC32;

Best Local Similarity 38.9%; Pred. No. 5.07e+00;  
Matches 14; Conservative 8; Mismatches 11; Indels 3; Gaps 2;

[illegible]

Query Match 30.6%; Score 71; DB 14; Length 351;  
Best Local Similarity 34.5%; Pred. No. 5.07e+00;  
Matches 10; Conservative 10; Mismatches 7; Indels

[illegible]

Query Match 30.6%; Score 71; DB 14; Length 357;  
Best Local Similarity 37.9%; Pred. No. 5.07e+00;  
Matches 11; Conservative 9; Mismatches 7; Indels 2; Gaps 2;

```

Db      259 FE-ATGNLVPR-YAFAMNRDPGSGIIIS 285
      : : : : : : : : : : : : : : : : : :
QY      65 FDPSTGALVDSKSYAFSTNSDNTTSAAFVS 93

```

Search completed: Wed Sep 1 16:24:42 1999

\*\*\*\*\*  
M P S R C H  
\*\*\*\*\* (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Sep 1 16:28:14 1999; MasPar time 3.49 Seconds  
Tabular output not generated. 48.772 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (57-64) from PCTUS9913024.pep (10 of 12)  
Perfect Score: 56  
Sequence: 1 SAGLHLRV 8

Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 15.625; Variance 42.633; scale 0.366

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	47	83.9	28	7_R34572	Domain 1 from integra	6.95e+01
2	45	80.4	185	32_W60046	Human TNF receptor re	1.20e+02
3	45	80.4	240	21_W05810	Human tumour necrosis	1.20e+02
4	44	78.6	467	36_W76426	Glucocorticoid induci	1.56e+02
5	44	78.6	467	36_W76427	Rat Glucocorticoid in	1.56e+02
6	44	78.6	843	3_R14924	Mutant SP6DNA polymer	1.56e+02
7	44	78.6	844	3_R14927	Mutant SP6DNA polymer	1.56e+02
8	44	78.6	849	3_R14922	SP6DNA polymerase	1.56e+02
9	44	78.6	849	3_R14923	Mutant SP6DNA polymer	1.56e+02
10	44	78.6	849	3_R14925	Mutant SP6DNA polymer	1.56e+02
11	44	78.6	849	3_R14926	Mutant SP6DNA polymer	1.56e+02
12	43	76.8	523	1_P80674	NtrA gene product.	2.04e+02
13	43	76.8	1891	26_W22610	Platenolide synthase	2.04e+02
14	43	76.8	1891	26_W23720	Platenolide synthase	2.04e+02
15	42	75.0	112	34_W77578	Staphylococcus aureus	2.66e+02
16	42	75.0	172	16_R82811	Rhhipcephalus appendi	2.66e+02

17	42	75.0	438	25	W18209	Staphylococcus aureus	2.66e+02
18	42	75.0	655	6	R31041	smrR polypeptide.	2.66e+02
19	42	75.0	1073	28	W32063	Human ST receptor pro	2.66e+02
20	42	75.0	1073	28	W37371	Human ST receptor pro	2.66e+02
21	42	75.0	1075	8	R38861	GC-C.	2.66e+02
22	42	75.0	1349	33	W59359	Human retinal degener	2.66e+02
23	42	75.0	1422	16	R82066	Hepatitis GB virus (H	2.66e+02
24	42	75.0	1706	16	R92103	Rat RIZ.	2.66e+02
25	41	73.2	30	37	W80945	Amino acid sequence o	3.45e+02
26	41	73.2	33	37	W80946	Amino acid sequence o	3.45e+02
27	41	73.2	108	18	W80946	G-CSF receptor ligand	3.45e+02
28	41	73.2	213	18	W00647	Mouse G-CSF receptor	3.45e+02
29	41	73.2	255	18	R99142	Ligand binding cytoki	3.45e+02
30	41	73.2	299	39	W80984	Amino acid sequence o	3.45e+02
31	41	73.2	299	39	W80985	B cell receptor assoc	3.45e+02
32	41	73.2	299	34	W61629	Clone HMWS46 of Proh	3.45e+02
33	41	73.2	334	18	R99140	Ligand binding cytoki	3.45e+02
34	41	73.2	412	37	W80942	Human heart lecithin-	3.45e+02
35	41	73.2	444	37	W80943	Human kidney lecithin-	3.45e+02
36	41	73.2	480	17	R85310	Cystathionine gamma s	3.45e+02
37	41	73.2	509	17	R85311	Cystathionine gamma s	3.45e+02
38	41	73.2	837	3	R14254	Murine GCSF receptor	3.45e+02
39	40	71.4	308	20	W10666	Human TGF-beta-like c	4.47e+02
40	40	71.4	308	20	W10670	Human TGF-beta-like c	4.47e+02
41	40	71.4	308	20	W10673	Human TGF-beta-like c	4.47e+02
42	40	71.4	308	16	R84710	Human embryonal lung	4.47e+02
43	40	71.4	323	1	R05711	Tumour-associated ant	4.47e+02
44	40	71.4	4303	17	R90302	Polycystic kidney dis	4.47e+02
45	40	71.4	4339	15	R75916	Polycystic kidney dis	4.47e+02

## ALIGNMENTS

RESULT 1  
ID R34572 standard; Protein; 28 AA.  
AC R34572;  
DE 04-JUN-1993 (first entry)  
DE Domain 1 from integrase like protein Tn2603 ORF3.  
KW cytotaxic T-lymphocyte response; transformed Mycobacteria; BCG;  
KW Mycobacterium smegmatis; vaccine; cell mediated immunity; HIV;  
KW pertussis; malaria; influenza virus; CTL; herpes virus.  
PN W09307897-A.  
PD 29-APR-1993.  
PF 21-OCT-1992; U09075.  
PR 21-OCT-1991; US-780261.  
PA (MEDI-) MEDIMUNE INC.  
PI Stover CK;  
DR WPI; 93-152187/18.  
PT Expression vector for expressing protein or polypeptide in  
PT mycobacterium - contg DNA sequences encoding lipoprotein  
PT secretion signal and peptide heterologous to bacteria expressing  
PT fusion protein of lipoprotein heterologous to bacteria  
PS Disclosure; Fig 29; 86pp; English.  
CC This sequence is domain 1 from the integrase like protein Tn2603 ORF 3.  
CC The whole coding sequence of the enzyme could be included in an  
CC expression vector in order to aid its integration into its host. The  
CC expression vector also includes at least the secretion signal of a  
CC lipoprotein and a second sequence encoding a heterologous protein  
CC and a mycobacterial promoter to control expression of the  
CC heterologous protein.  
SQ Sequence 28 AA;  
Query Match 83.9%; Score 47; DB 7; Length 28;  
Best Local Similarity 75.0%; Pred. No. 6.95e+01;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Db 15 segqlgrv 22  
| | | | |  
QY 57 SAGLHLRV 64  
RESULT 2  
ID W60046 standard; Protein; 185 AA.



AC W60046;  
DT 25-SEP-1998 (first entry)  
DE Human TNF receptor related splice variant 1 (TR2-SV1) protein.  
KW Human: tumour necrosis factor; TNF; herpes simplex virus;  
KW aberrant cell survival; radiation therapy; lymphocyte proliferation;  
KW immune deficiency syndrome.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Peptide 1..36  
FT /note= "Signal peptide"  
FT 37..185  
FT /note= "Mature peptide"  
FT W09818824-A1.  
PN 07-MAY-1998.  
PD 30-OCT-1996; U18540.  
PR 30-OCT-1996; WO-U18540.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PI Gentz RL, Hurler MR, Lyn SDP, Ni J, Rosen CA;  
DR WPI: 98-272139/24.  
DR N-PSDB; V34510.  
DR Nucleic acid encoding TR2 tumour necrosis factor family receptor -  
PT and its splice variants, useful for diagnosis and treatment of  
PT diseases involving abnormal cell survival or death, e.g. herpes  
PT simplex infection  
PS Claim 1: Fig 4: 151pp; English.  
CC The human tumour necrosis factor (TNF) receptor related protein (TR2) is  
CC a member of the TNF family and displays considerable homology to murine  
CC CD40. It can be used in soluble forms to treat herpes simplex virus  
CC infection and TR2 proteins (or their agonists or antagonists) are used to  
CC treat disease associated with aberrant cell survival. Agonists may also  
CC be used to protect against the effects of radiation therapy and to  
CC stimulate lymphocyte proliferation and differentiation in patients  
CC infected by human immune deficiency syndrome.  
SQ Sequence 185 AA;  
  
Query Match 80.4%; Score 45; DB 32; Length 185;  
Best Local Similarity 100.0%; Pred. No. 1.20e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 169 aglhr 174  
| | | | |  
QY 58 AGLHLR 63  
  
RESULT 3  
ID W05810 standard; Protein; 240 AA.  
AC W05810;  
DT 30-JUN-1997 (first entry)  
DE Human tumour necrosis factor receptor splice variant 1.  
KW Human: tumour necrosis factor; TNF; receptor; treatment;  
KW activation; inhibition; identification; agonist; antagonist;  
KW stimulation; T cell; differentiation; mediation; immune;  
KW antiviral; response; regulation; growth; protection; HIV;  
KW radiation; Chlamydia; infection; immunodeficiency; virus;  
KW autoimmune disease; inflammation; septic shock; cerebral; malaria;  
KW cachexia; B cell; cancer; graft; host; reaction; rejection;  
KW prevention; apoptosis; cytotoxicity; hybridisation probe;  
KW detection; antibody; reagent; splice variant 1; SV1.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Peptide 1..38  
FT /label= sig\_peptide  
FT 39..240  
FT /label= mat\_peptide  
FT W09634095-A1.  
PN 31-OCT-1996.  
PD 27-APR-1995; U05058.  
PR 27-APR-1995; WO-U05058.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Gentz R, Ni J, Rosen CA;  
DR WPI: 96-497627/49.  
DR N-PSDB; T40049.

PT New nucleic acid encoding a human tumour necrosis factor receptor -  
PT useful for treatment of auto-immune diseases etc., in diagnosis and  
PT for drug screening  
PS Claim 1: Pages 48-49: 73pp; English.  
CC The present sequence is the human tumour necrosis factor (TNF)  
CC receptor splice variant 1 (SV1), which may be used for the  
CC identification of TNF SV1 receptor agonists or antagonists. TNF SV1  
CC receptor agonists inhibit tumour growth, stimulate cell (e.g.  
CC T cell) differentiation, mediate the immune and antiviral  
CC responses, regulate growth, protect against radiation and Chlamydia  
CC infection, and can be used to treat immunodeficiencies, e.g. human  
CC immunodeficiency virus. Antagonists can be used to treat T cell  
CC mediated autoimmune disease, inflammation, septic shock, cerebral  
CC malaria, cachexia or B cell cancers, to inhibit graft-host  
CC reactions and to prevent apoptosis or cytotoxicity. Fragments of  
CC the TNF SV1 receptor encoding DNA can be used as hybridisation  
CC probes for detecting related genes. Antibodies against the protein  
CC can be used as reagents for detecting/measuring soluble forms of  
CC protein in the circulation.  
SQ Sequence 240 AA;  
  
Query Match 80.4%; Score 45; DB 21; Length 240;  
Best Local Similarity 100.0%; Pred. No. 1.20e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 169 aglhr 174  
| | | | |  
QY 58 AGLHLR 63  
  
RESULT 4  
ID W76426 standard; Protein; 467 AA.  
AC W76426;  
DT 15-JAN-1999 (first entry)  
DE Glucocorticoid inducible suppressor protein.  
KW Glucocorticoid inducible suppressor protein; interleukin-8; rheumatism;  
KW intracellular signal transduction inhibitor; IL-8 promoter; allergy;  
KW inflammatory disease; bronchial asthma; therapy.  
OS Homo sapiens.  
PN W09838213-A1.  
PD 03-SEP-1998.  
PF 27-FEB-1998; J00836.  
PR 28-FEB-1997; JP-062008.  
PA (CYTO-) INST CYTOSIGNAL RES INC.  
PI Ohtsuka T, Yoshida H;  
DR WPI: 98-481142/41.  
DR N-PSDB; V61810.  
DR Protein inhibiting activation of interleukin-8 promoter in response  
PT to extracellular stimulus - for use in diseases involving  
PT interleukin-8 expression, such as inflammatory disorders and asthma  
PS Claim 1: Page 39-43: 72pp; Japanese.  
CC This sequence represents the human glucocorticoid inducible suppressor  
CC protein (GISP) of the invention. The GISP is an intracellular signal  
CC transduction inhibitor, and inhibits the activation of interleukin-8  
CC (IL-8) promoter activity in response to a specific extracellular stimulus  
CC (especially by interleukin-1 beta). The invention can be used in the  
CC treatment and prevention of disorders in which IL-8 expression is  
CC involved, such as inflammatory diseases, bronchial asthma, allergy and  
CC rheumatism.  
SQ Sequence 467 AA;  
  
Query Match 78.6%; Score 44; DB 36; Length 467;  
Best Local Similarity 71.4%; Pred. No. 1.56e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Db 32 aglhr 38  
| | | | |  
QY 57 SAGLHLR 63  
  
RESULT 5  
ID W76427 standard; Protein; 467 AA.  
AC W76427;

DT 15-JAN-1999 (first entry)  
 DE Rat Glucocorticoid inducible suppressor protein.  
 KW Glucocorticoid inducible suppressor protein; Interleukin-8; rheumatism;  
 KW Intracellular signal transmission inhibitor; IL-8 promoter; allergy;  
 KW Inflammatory disease; bronchial asthma; therapy.  
 OS Rattus sp.  
 PN WO9838213-A1.  
 PD 03-SEP-1998.  
 PF 27-FEB-1998; J00836.  
 PR 28-FEB-1997; JP-062008.  
 PA (CYTO-) INST CYTOSIGNAL RES INC.  
 PI Ohtsuka T, Yoshida H,  
 DR WPI; 98-481142/41.  
 DR N-PSDB; V61811.  
 PT Protein inhibiting activation of interleukin-8 promoter in response  
 PT to extracellular stimulus - for use in diseases involving  
 PT interleukin-8 expression, such as inflammatory disorders and asthma  
 PS Claim 3; Page 46-50; 72pp; Japanese.  
 CC This sequence represents the rat glucocorticoid inducible suppressor  
 CC protein (GISP) of the invention. The GISP is a intracellular signal  
 CC transmission inhibitor, and inhibits the activation of interleukin-8  
 CC (IL-8) promoter activity in response to a specific extracellular stimulus  
 CC (especially by interleukin-1 beta). The invention can be used in the  
 CC treatment and prevention of disorders in which IL-8 expression is  
 CC involved, such as inflammatory diseases, bronchial asthma, allergy and  
 CC rheumatism.  
 SQ Sequence 467 AA;

Query Match 78.6%; Score 44; DB 36; Length 467;  
 Best Local Similarity 71.4%; Pred. No. 1.56e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 31 apghlhr 37  
 :|||||  
 QY 57 SAGLHLRV 63

## RESULT 6

ID R14924 standard; Protein; 843 AA.  
 AC R14924;  
 DT 11-FEB-1992 (first entry)  
 DE Mutant SP6DNA polymerase.  
 KW Exonuclease.  
 OS SP6 phage.  
 PN J03232487-A.  
 PD 16-OCT-1991.  
 PF 06-FEB-1990; 025185.  
 PR 06-FEB-1990; JP-025185.  
 PA (TAKA-) TAKARA SHUZO KK.  
 DR WPI; 91-349003/48.  
 PT SP6DNA polymerase of defined sequence - and modified polymerase  
 PT lacking 3',5'-exo-nuclease activity  
 PS Disclosure; Page 8; 13pp; Japanese  
 CC The amino acid sequence is that of mutated SP6DNA polymerase, where  
 CC amino acids His167-His172 have been deleted by site specific  
 CC mutation. This results in a mutant lacking 3',5'-exonuclease  
 CC activity. See also R14922-R14927.  
 SQ Sequence 843 AA;

Query Match 78.6%; Score 44; DB 3; Length 843;  
 Best Local Similarity 75.0%; Pred. No. 1.56e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 565 gaglelrv 572  
 :|||||  
 QY 57 SAGLHLRV 64

## RESULT 7

ID R14927 standard; Protein; 844 AA.  
 AC R14927;  
 DT 11-FEB-1992 (first entry)  
 DE Mutant SP6DNA polymerase.

KW Exonuclease.  
 OS SP6 phage.  
 PN J03232487-A.  
 PD 16-OCT-1991.  
 PF 06-FEB-1990; 025185.  
 PR 06-FEB-1990; JP-025185.  
 PA (TAKA-) TAKARA SHUZO KK.  
 DR WPI; 91-349003/48.  
 PT SP6DNA polymerase of defined sequence - and modified polymerase  
 PT lacking 3',5'-exo-nuclease activity  
 PS Disclosure; Page 8; 13pp; Japanese  
 CC The amino acid sequence is that of mutated SP6DNA polymerase, where  
 CC amino acids Val164-Ser168 have been deleted by site specific  
 CC mutation. See also R14922-R14926.  
 SQ Sequence 844 AA;

Query Match 78.6%; Score 44; DB 3; Length 844;  
 Best Local Similarity 75.0%; Pred. No. 1.56e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 566 gaglelrv 573  
 :|||||  
 QY 57 SAGLHLRV 64

## RESULT 8

ID R14922 standard; Protein; 849 AA.  
 AC R14922;  
 DT 11-FEB-1992 (first entry)  
 DE SP6DNA polymerase.  
 KW Exonuclease.  
 OS SP6 phage.  
 PN J03232487-A.  
 PD 16-OCT-1991.  
 PF 06-FEB-1990; 025185.  
 PR 06-FEB-1990; JP-025185.  
 PA (TAKA-) TAKARA SHUZO KK.  
 DR WPI; 91-349003/48.  
 DR N-PSDB; Q14772.  
 PT SP6DNA polymerase of defined sequence - and modified polymerase  
 PT lacking 3',5'-exo-nuclease activity  
 PS Claim 1; Fig 1; 13pp; Japanese  
 CC The amino acid sequence is that of SP6DNA polymerase and was obtd.  
 CC from the SP6 phage. It can be produced recombinantly and is of  
 CC property, superior to conventional polymerases. See also R14923-  
 CC R14927  
 SQ Sequence 849 AA;

Query Match 78.6%; Score 44; DB 3; Length 849;  
 Best Local Similarity 75.0%; Pred. No. 1.56e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 571 gaglelrv 578  
 :|||||  
 QY 57 SAGLHLRV 64

## RESULT 9

ID R14923 standard; Protein; 849 AA.  
 AC R14923;  
 DT 11-FEB-1992 (first entry)  
 DE Mutant SP6DNA polymerase.  
 KW Exonuclease.  
 OS SP6 phage.  
 PN J03232487-A.  
 PD 16-OCT-1991.  
 PF 06-FEB-1990; 025185.  
 PR 06-FEB-1990; JP-025185.  
 PA (TAKA-) TAKARA SHUZO KK.  
 DR WPI; 91-349003/48.  
 PT SP6DNA polymerase of defined sequence - and modified polymerase  
 PT lacking 3',5'-exo-nuclease activity  
 PS Disclosure; Page 8; 13pp; Japanese

CC The amino acid sequence is that of mutated SP6DNA polymerase, where  
CC the Asp at position 101 has been changed by site specific mutation  
CC to Ala. See also R14922-R14927.  
SQ Sequence 849 AA;

Query Match 78.6%; Score 44; DB 3; Length 849;  
Best Local Similarity 75.0%; Pred. No. 1.56e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 571 gaglelrv 578  
:|||||  
QY 57 SAGLHLRV 64

RESULT 10  
ID R14925 standard; Protein; 849 AA.

AC R14925.

DT 11-FEB-1992 (first entry)

DE Mutant SP6DNA polymerase.

KW Exonuclease.

OS SP6 phage.

PN J03232487-A.

PD 16-OCT-1991.

PF 06-FEB-1990; 025185.

PR 06-FEB-1990; JP-025185.

PA (TAKA-) TAKARA SHUZO KK.

WPI: 91-349003/48.

PT SP6DNA polymerase of defined sequence - and modified polymerase

PT lacking 3', 5'-exo-nuclease activity

PS Disclosure; Page 8; 13pp; Japanese

CC The amino acid sequence is that of mutated SP6DNA polymerase, where

CC the His at position 167 has been changed by site specific mutation

CC to Glu, Glu170 to Lys and His172 to Asp. See also R14922-R14927.

SQ Sequence 849 AA;

Query Match 78.6%; Score 44; DB 3; Length 849;  
Best Local Similarity 75.0%; Pred. No. 1.56e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 571 gaglelrv 578  
:|||||  
QY 57 SAGLHLRV 64

RESULT 11  
ID R14926 standard; Protein; 849 AA.

AC R14926.

DT 11-FEB-1992 (first entry)

DE Mutant SP6DNA polymerase.

KW Exonuclease.

OS SP6 phage.

PN J03232487-A.

PD 16-OCT-1991.

PF 06-FEB-1990; 025185.

PR 06-FEB-1990; JP-025185.

PA (TAKA-) TAKARA SHUZO KK.

WPI: 91-349003/48.

PT SP6DNA polymerase of defined sequence - and modified polymerase

PT lacking 3', 5'-exo-nuclease activity

PS Disclosure; Page 8; 13pp; Japanese

CC The amino acid sequence is that of mutated SP6DNA polymerase, where

CC the His at position 167 has been changed by site specific mutation

CC to Gln and Ser168 to Asn. See also R14922-R14927.

SQ Sequence 849 AA;

Query Match 78.6%; Score 44; DB 3; Length 849;  
Best Local Similarity 75.0%; Pred. No. 1.56e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 571 gaglelrv 578  
:|||||  
QY 57 SAGLHLRV 64

RESULT 12

ID P80674 standard; Protein; 523 AA.

AC P80674;

DT 24-OCT-1990 (first entry)

DE NtrA gene product.

KW NtrA; sigma factor; transcriptional activation; nitrogen assimilation;

KW fixation; nif operon; C4-dicarboxylate transport.

OS Rhizobium meliloti.

PN EP-292984-A.

PD 30-NOV-1988.

PF 27-MAY-1988; 108482.

PR 29-MAY-1987; US-055228.

PA (GEHO-) Gen Hospital Corp.

PI Ronson C, Ausubel F;

WPI: 88-339561/48.

DR N-PSDB; N81266.

PT Cloned rhizobium meliloti ntrA gene - for use in C4-dicarboxylate

PT transport, nitrate assimilation, symbiotic nitrogen fixation and

PS identifying pathogenic genes.

PS Claim 1; Page 12; 19pp; English.

CC The ntrA gene product is a transcriptional activator. It controls the

CC processes of nitrate assimilation, symbiotic nitrogen fixation and

CC C4-dicarboxylate transport, working in conjunction with NtrC, NifA and

CC DctD resp. It can also be used to identify genes involved in

CC pathogenesis.

CC See also P82126.

SQ Sequence 523 AA;

Query Match 76.8%; Score 43; DB 1; Length 523;  
Best Local Similarity 85.7%; Pred. No. 2.04e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 4 sasllhr 10  
|||||  
QY 57 SAGLHLR 63

RESULT 13

ID W22610 standard; Protein; 1891 AA.

AC W22610;

DT 27-FEB-1998 (first entry)

DE Platenolide synthase ORF5 protein.

KW Tyactone synthase gene cluster; tyLG gene; multifunctional protein;

KW polyketide; tyactone synthesis; antibiotic; tylosin.

OS Streptomyces ambofaciens.

PH Key Location/Qualifiers

FT Domain 35..458

FT Domain /note= "ketosynthase domain, KS7"

FT Domain 582..917

FT Domain /note= "acyltransferase domain, AT7"

FT Domain 1233..1418

FT Domain /note= "ketoreductase domain, KR7"

FT Domain 1502..1585

FT Domain /note= "acyl carrier protein domain, ACP7"

FT Domain 1715..1747

FT Domain /note= "thioesterase domain, TE7"

PN EP-791655-A2.

PD 27-AUG-1997.

PF 19-FEB-1997; 301056.

PR 22-FEB-1996; US-012078.

PA (EJIL) LILLY & CO ELI.

PI Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;

WPI: 97-418046/39.

DR N-PSDB; T80414.

DR DNA encoding Streptomyces fradiae tyactone synthase domain - for

PT production of tylosin-related polyketide compounds

PS Example 2; Pages 172-178; 220pp; English.

CC W22606-W22610 represent proteins encoded by the platenolide synthase gene

CC cluster. The gene cluster is also referred to as the srmg gene, and was

CC isolated from Streptomyces ambofaciens. These sequences are

CC multi-functional proteins which direct the synthesis of the polyketide

CC platenolide. Platenolide is the basic building block of the macrolide

CC antibiotic spiramycin. The DNA encoding this sequence was used along with  
 CC the tylg gene (see T80413) to create a hybrid ORF1 sequence (see T80415).  
 CC The tylg gene is the tyllactone synthase gene cluster of the invention.  
 CC The tylg sequence was isolated from Streptomyces fradiae, and encodes  
 CC multifunctional proteins which direct the synthesis of the polyketide  
 CC tyllactone. Tyllactone is the basic building block of the antibiotic  
 CC tylosin. The hybrid sequence can be used to transform S. ambofaciens  
 CC lacking the srmg ORF1 sequence, or S. fradiae lacking the tylg ORF1  
 CC sequence, so that they can produce polyketides. The DNA sequence can be  
 CC modified so as to alter the type of carboxylic acids incorporated, the  
 CC number of carboxylic acids incorporated and/or the post-condensation  
 CC reactions performed, thereby resulting in novel tylosin-related  
 CC polyketides.  
 CC Sequence 1891 AA;

Query Match 76.8%; Score 43; DB 26; Length 1891;  
 Best Local Similarity 75.0%; Pred. No. 2.04e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 1007 taglhvrv 1014  
 :|||||  
 QY 57 SAGLHLRV 64

RESULT 14  
 ID W23720 standard; Protein; 1891 AA.  
 AC W23720;  
 DT 27-FEB-1998 (first entry)  
 DE Platenolide synthase ORF5 protein.  
 KW Platenolide synthase gene cluster; platenolide production; srmg gene;  
 KW multi-functional protein; macrolide antibiotic; spiramycin.  
 OS Streptomyces ambofaciens.  
 FH Key Location/Qualifiers  
 FT Domain 35..458  
 FT Domain 582..917  
 FT Domain 1235..1418  
 FT Domain /note= "acyltransferase domain, A77"  
 FT Domain /note= "ketoreductase domain, KR7"  
 FT Domain 1502..1585  
 FT Domain /note= "acyl carrier protein domain, ACP7"  
 FT Domain 1715..1747  
 FT Domain /note= "thioesterase domain, TE7"  
 PN EP-791656-A2.  
 PD 27-AUG-1997.  
 PF 19-FEB-1997; 301066.  
 PR 22-FEB-1996; US-012050.  
 PA (ELIL) LILLY & CO ELI.  
 PI Burgett SG, Kuhstoss SA, Rao RN, Richardson MA;  
 PI Rosteck PR;  
 DR WPI; 97-418047/39.  
 DR N-PSDB; T78508.  
 PT DNA encoding Streptomyces ambofaciens platenolide synthase domain -  
 PT for production of spiramycin-related polyketide antibiotics  
 PS Claim 8; Pages 71-77; 81pp; English.  
 CC W23716-W23720 represent proteins encoded by the platenolide synthase gene  
 CC cluster of the invention. The gene cluster is also referred to as the  
 CC srmg gene, and was isolated from Streptomyces ambofaciens. These  
 CC sequences are multi-functional proteins which direct the synthesis of the  
 CC polyketide platenolide. Platenolide is the basic building block of the  
 CC macrolide antibiotic spiramycin. The DNA can be used to produce compounds  
 CC exhibiting antibiotic activity based on the platenolide structure,  
 CC including specifically the macrolide antibiotic spiramycin and spiramycin  
 CC analogues and derivatives. Modifications of the platenolide synthase DNA  
 CC sequence can be made so as to change the number and type of carboxylic  
 CC acids incorporated into the growing polyketide chain and to change the  
 CC kind of post-condensation processing that is conducted.  
 CC Sequence 1891 AA;

Query Match 76.8%; Score 43; DB 26; Length 1891;  
 Best Local Similarity 75.0%; Pred. No. 2.04e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 1007 taglhvrv 1014  
 :|||||  
 QY 57 SAGLHLRV 64

RESULT 15  
 ID W77578 standard; Protein; 112 AA.  
 AC W77578;  
 DT 30-OCT-1998 (first entry)  
 DE Staphylococcus aureus protein of unknown function.  
 KW Staphylococcus aureus protein; immune response induction; eye infection;  
 KW antibody production; T-cell immune response; gastrointestinal infection;  
 KW respiratory infection; inhibitor; bacterial infection; cardiac infection;  
 KW central nervous system; kidney infection; urinary tract infection;  
 KW antimicrobial compound identification; broad spectrum antibiotic;  
 KW therapy.  
 OS Staphylococcus aureus.  
 FH Key Location/Qualifiers  
 FT Misc\_difference 1..112  
 FT /note= "residues designated X are unspecified, and  
 FT represented as Xaa in the specification"

EP-841394-A2.  
 13-MAY-1998.  
 PD 24-SEP-1997; 307485.  
 PF 24-SEP-1996; US-027032.  
 PR (SMIK) SMITHKLINE BEECHAM CORP.  
 PA (SMIK) SMITHKLINE BEECHAM PLC.  
 PI Black MT, Burnham MKR, Hodgson JE, Knowles DJC,  
 PI Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,  
 PI Ward JM;  
 DR WPI; 98-252940/23.  
 DR N-PSDB; V53376.  
 PT New nucleic acid sequences from Staphylococcus aureus WCHU29 -  
 PT useful in vaccines and for treatment of bacterial infections of e.g.  
 PT respiratory tract and central nervous system  
 PS Claim 11; Page 263; 390pp; English.  
 CC This sequence represents a Staphylococcus aureus protein of unknown  
 CC function, and is encoded by a DNA sequence of the invention.  
 CC The DNA sequences were isolated from Staphylococcus aureus WCHU29  
 CC (NCIMB 40771). Host cells containing the DNA sequences are used to  
 CC produce polypeptides or fragments. The proteins are used in the treatment  
 CC of disease, for inducing an immune response by administering them, to  
 CC produce antibody and/or T-cell immune response. Antagonists of the  
 CC proteins are used for the inhibition of bacterial polypeptides.  
 CC Conditions which may be treated include bacterial infections, especially  
 CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,  
 CC urinary tract, skin, bones and joints. The proteins can also be used to  
 CC identify antimicrobial compounds which are broad spectrum antibiotics,  
 CC especially useful in the treatment of H. pylori infection.  
 CC Sequence 112 AA;

Query Match 75.0%; Score 42; DB 34; Length 112;  
 Best Local Similarity 62.5%; Pred. No. 2.66e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 40 nsp1hvr 47  
 :|||||  
 QY 57 SAGLHLRV 64

Search completed: Wed Sep 1 16:28:39 1999  
 Job time : 25 secs.

\*\*\*\*\*  
 W A S E R L A  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Wed Sep 1 16:27:45 1999; MasPar time 3.59 Seconds  
 Tabular output not generated.  
 89.384 Million cell updates/sec

Title: >PCT-US99-13024-2  
 Description: (57-64) from PCTUS9913024.pap (10 of 12)  
 Perfect Score: 56  
 Sequence: 1 SAGLHLRV 8

Scoring table: PAM 150  
 Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: pir60  
 1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 21.865; Variance 24.608; scale 0.889

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description	Pred. No.
1	48	85.7	415	2	T02145	hypothetical protein
2	47	83.9	303	2	S10927	integrase-like protei
3	47	83.9	337	2	C26839	integrase - Escherich
4	47	83.9	337	2	I39499	integrase - Acinetoba
5	47	83.9	337	2	S32184	integrase - Klebsiell
6	47	83.9	337	2	A42646	integrase - Pseudomon
7	47	83.9	337	2	J00300	hypothetical 38.4K pr
8	45	80.4	499	2	F69833	monooxygenase homolog
9	44	78.6	147	2	D27733	hypothetical protein
10	44	78.6	461	2	S52408	SPDVR1 protein - sea
11	44	78.6	572	1	VEHULC	lamin C - human
12	44	78.6	574	2	S04333	lamin C - mouse
13	44	78.6	657	2	S05517	lamin - chicken
14	44	78.6	664	1	VEHULA	lamin A - human
15	44	78.6	665	2	S28182	lamin A - mouse
16	44	78.6	665	2	S27267	lamin A - rat
17	43	76.8	341	1	H69401	probable radical-form
18	43	76.8	373	2	A27817	lignin peroxidase (EC
19	43	76.8	513	2	A33966	ntrA protein - Rhizob
20	43	76.8	525	2	A36130	transcription initiat
21	43	76.8	1117	2	S38673	desmoglein 2 - human
22	42	75.0	211	2	S65049	low molecular weight
23	42	75.0	218	2	T01412	low molecular weight

24	42	75.0	325	2	JN0148	neccin, brain - mouse	5.97e+01
25	42	75.0	349	2	E70640	hypothetical protein	5.97e+01
26	42	75.0	354	2	JE0309	ornithine carbamoyltr	5.97e+01
27	42	75.0	380	2	T01706	hypothetical protein	5.97e+01
28	42	75.0	461	2	S54543	probable membrane pro	5.97e+01
29	42	75.0	489	2	S69027	ammonium transport pr	5.97e+01
30	42	75.0	494	2	S62902	legumin 2 precursor -	5.97e+01
31	42	75.0	511	2	A40110	nicotinic acetylcholi	5.97e+01
32	42	75.0	524	2	F70579	probable cell divisio	5.97e+01
33	42	75.0	604	2	S25203	smr protein - Strept	5.97e+01
34	42	75.0	627	1	BHTLD	hemocyanin chain d -	5.97e+01
35	42	75.0	698	2	A47203	protein-glutamine gam	5.97e+01
36	42	75.0	702	2	C54835	hypothetical protein	5.97e+01
37	42	75.0	730	2	A70680	hypothetical protein	5.97e+01
38	42	75.0	818	2	JC4397	peroxinectin precurs	5.97e+01
39	42	75.0	862	2	D70199	DNA mismatch repair p	5.97e+01
40	42	75.0	1073	1	OYRHX	heat-stable enterotox	5.97e+01
41	42	75.0	1075	1	OYRTHX	heat-stable enterotox	5.97e+01
42	42	75.0	1287	2	A41685	SIL protein - human	5.97e+01
43	42	75.0	1706	2	T84499	zinc finger protein R	5.97e+01
44	41	73.2	509	3	T02942	O-succinylhomoserine	9.49e+01
45	41	73.2	5262	2	T03454	ALR protein - human	9.49e+01

## ALIGNMENTS

RESULT 1  
 ENTRY T02145 #type complete  
 TITLE hypothetical protein F8K4.20 - Arabidopsis thaliana  
 ORGANISM #formal\_name Arabidopsis thaliana #common\_name mouse-ear cress  
 DATE 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 05-Mar-1999  
 ACCESSIONS T02145  
 REFERENCE T14574  
 #authors Vysotskaya, V.S.; Schwartz, J.R.; Toriumi, M.; Kwan, A.; Yu, G.; Oji, O.; Liu, S.; Li, J.; Araujo, R.; Au, M.; Brendel, V.; Buehler, E.; Conway, A.B.; Conway, A.R.; Dewar, K.; Feng, J.; Kim, C.; Kurtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel, N.A.; Theologis, A.  
 #submission submitted to the EMBL Data Library, August 1998  
 #description Arabidopsis thaliana chromosome 1 BAC F8K4 sequence.  
 #accession T02145  
 #status preliminary; translated from GB/EMBL/DBJ  
 ##molecule\_type DNA  
 ##residues 1-415 ##label VYS  
 ##cross-references EMBL:AC004392; NID:g3282170; PID:g3367531

GENETICS  
 #map\_position 1  
 #note F8K4.20  
 SUMMARY #length 415 #molecular-weight 46438 #checksum 9304  
 Query Match 85.7%; Score 48; DB 2; Length 415;  
 Best Local Similarity 100.0%; Pred. No. 3.06e+00;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 14 GLHLRV 19  
 QY 59 GLHLRV 64

RESULT 2  
 ENTRY S10927 #type complete  
 TITLE integrase-like protein 2M - Mycobacterium fortuitum  
 ORGANISM #formal\_name Mycobacterium fortuitum  
 DATE 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 08-Sep-1997  
 ACCESSIONS S10927  
 REFERENCE S10927  
 #authors Martin, C.; Timm, J.; Rauszier, J.; Gomez-Lus, R.; Davies, J.; Gicquel, B.

```

#journal      Nature (1990) 345:739-743
#title        Transposition of an antibiotic resistance element in
#              mycobacteria.
#cross-references EMBL:X53635; NID:g44283; PID:g44286
#accession     S10927
#molecule_type DNA
#residues      1-303 ##label NAT
##cross-references EMBL:X53635; NID:g44283; PID:g44286
CLASSIFICATION #superfamily integrase-like protein
SUMMARY        #length 303 #molecular-weight 34884 #checksum 5690

Query Match      83.9%; Score 47; DB 2; Length 303;
Best Local Similarity 75.0%; Pred. NO. 5.12e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155
QY 57 SAGLHLRV 64

RESULT 3
ENTRY C26839 #type complete
TITLE integrase - Escherichia coli plasmid and transposons
ALTERNATE_NAMES recombinase
ORGANISM #formal_name Escherichia coli
DATE 20-Jun-1989 #sequence_revision 20-Jun-1989 #text_change
13-Mar-1998
ACCESSIONS C26839; S21844; S31856; S04809; I60356; S60664
REFERENCE A26839
#authors Hall, R.M.; Vockler, C.
#journal Nucleic Acids Res. (1987) 15:7491-7501
#title The region of the IncN plasmid R46 coding for resistance to
beta-lactam antibiotics, streptomycin/spectinomycin and
sulphonamides is closely related to antibiotic resistance
segments found in IncW plasmids and in Tn21-like
transposons
#cross-references MUID:88015610
#accession C26839
##molecule_type DNA
##residues 1-337 ##label HAL
##cross-references GB:X06046; NID:g42204; PID:g42207
##experimental_source plasmid R46
#genetics PLA
REFERENCE S21844
#authors Sundstroem, L.; Swedberg, G.; Skold, O.
#submission Submitted to the EMBL Data Library, March 1991
#description Characterization of the transposon Tn5086, and its
site-specifically inserted gene dhfrVII, which encodes a
new type of trimethoprim-resistant dihydrofolate reductase.
#accession S21844
##status preliminary
##molecule_type DNA
##residues 1-145 ##label SUN
##cross-references EMBL:X58425; NID:g43090; PID:g43091
##experimental_source transposon Tn5086
#genetics TRA
REFERENCE S31854
#authors Heikkila, E.; Skurnik, M.; Sundstrom, L.; Huovinen, P.
#submission Submitted to the EMBL Data Library, February 1993
#description Another dihydrofolate reductase cassette, inserted in an
integron, borne on a Tn21-like element.
#accession S31856
##status preliminary
##molecule_type DNA
##residues 1-55 ##label HEI
##cross-references EMBL:221672; NID:g49350; PID:g49353
##experimental_source Tn21-like element
#genetics TLE
REFERENCE S04809
#authors Sundstroem, L.; Radstrom, P.; Swedberg, G.; Skoeld, O.
#journal Mol. Gen. Genet. (1988) 213:191-201
#title Site-specific recombination promotes linkage between
trimethoprim- and sulfonamide resistance genes. Sequence

```

```

characterization of dhfrv and sulI and a recombination
active locus of Tn21.
#cross-references MUID:89039710
#accession S04809
##status translation not shown
##molecule_type DNA
##residues 1-81,'p',83,'p',85-337 ##label SU2
##cross-references EMBL:X12868; NID:g45672; PID:g45673
##experimental_source plasmid pLMO20, Tn21
#genetics T21
REFERENCE I60356
#authors Radstrom, P.; Skold, O.; Swedberg, G.; Flensburg, J.; Roy,
P.H.; Sundstrom, L.
#journal J. Bacteriol. (1994) 176:3257-3268
#title Transposon Tn5090 of plasmid R751, which carries an integron,
is related to Tn7, Mu, and the retroelements.
#cross-references MUID:94252994
#accession I60356
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-145 ##label RES
##cross-references EMBL:X58425; NID:g43090; PID:g43091
##experimental_source plasmid R751, Tn5090
#genetics T50
REFERENCE S60664
#authors Adrian, P.V.; Thomson, C.J.; Klugman, K.P.; Ames, S.G.
#submission submitted to the EMBL Data Library, August 1995
#description Nucleotide sequence of the cassette-borne type XIII
dihydrofolate reductase from an Escherichia coli strain
isolated in South Africa.
#accession S60664
##status preliminary
##molecule_type DNA
##residues 1-191 ##label ADR
##cross-references EMBL:X50802; NID:g1841698; PID:g950043
##experimental_source strain isolated in South Africa
GENETICS PLA
#gene int
#genome int
GENETICS TRA
#gene int
#genome int
GENETICS TLE
#gene int
#genome int
GENETICS T21
#gene int
#genome int
GENETICS T50
#gene int
#genome int
CLASSIFICATION #superfamily integrase-like protein
SUMMARY #length 337 #molecular-weight 38381 #checksum 9946

Query Match      83.9%; Score 47; DB 2; Length 337;
Best Local Similarity 75.0%; Pred. NO. 5.12e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155
QY 57 SAGLHLRV 64

RESULT 4
ENTRY I39499 #type complete
TITLE Integrase - Acinetobacter baumannii
ORGANISM #formal_name Acinetobacter baumannii
DATE 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change
16-Feb-1997
ACCESSIONS I39499
REFERENCE I39499
#authors Elisha, B.G.; Steyn, L.M.
#journal Curr. Microbiol. (1991) 22:259-263

```

```

#title      Cloning of AAC(3) and AAD(2') genes from Acinetobacter:
#accession  I39499
##status    preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues  1-337 ##label RES
##cross-references EMBL:U17586; NID:g596249; PID:g596251
GENETICS
#gene       tnpl
CLASSIFICATION
#superfamily integrase-like protein
SUMMARY     #length 337 #molecular-weight 38381 #checksum 9946

Query Match      83.9%; Score 47; DB 2; Length 337;
Best Local Similarity 75.0%; Pred. No. 5.12e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155
| ||:|
QY 57 SAGLHLRV 64

RESULT 5
ENTRY
TITLE      #type complete
#formal_name Klebsiella pneumoniae
ORGANISM    integrase - Klebsiella pneumoniae
DATE        #formal_name Klebsiella pneumoniae
06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
08-Sep-1997
ACCESSIONS  S32184
REFERENCE     S32177
#authors     Radstroem, P.; Sundstrom, L.; Swedberg, G.; Flensburg, J.; Skold, O.
#submission  Submitted to the EMBL Data Library, March 1993
#description The integrons of R751 and Tn21 are transposons related to the
retroelements and to Tn7.
#accession   S32184
##status     preliminary
##molecule_type DNA
##residues   1-337 ##label RAD
##cross-references EMBL:X72585; NID:g288626; PID:g288634
CLASSIFICATION
#superfamily integrase-like protein
SUMMARY     #length 337 #molecular-weight 38358 #checksum 180

Query Match      83.9%; Score 47; DB 2; Length 337;
Best Local Similarity 75.0%; Pred. No. 5.12e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155
| ||:|
QY 57 SAGLHLRV 64

RESULT 6
ENTRY
TITLE      #type complete
#formal_name Pseudomonas aeruginosa
ORGANISM    integrase - Pseudomonas aeruginosa
DATE        #formal_name Pseudomonas aeruginosa
04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
12-Apr-1995
ACCESSIONS  A42646
REFERENCE     A42646
#authors     Bissonnette, L.; Roy, P.H.
#journal     J. Bacteriol. (1992) 174:1248-1257
#title       Characterization of in0 of Pseudomonas aeruginosa plasmid
pvsl, an ancestor of integrons of multiresistance plasmids
and transposons of gram-negative bacteria.
#cross-references MUID:92138617
#contents     plasmid pvsl
#accession   A42646
##status     preliminary
##molecule_type DNA
##residues   1-337 ##label BIS
##note        sequence extracted from NCBI backbone (NCBIN:79394,
NCBIP:79395)
CLASSIFICATION
#superfamily integrase-like protein

```

```

SUMMARY     #length 337 #molecular-weight 38381 #checksum 9946

Query Match      83.9%; Score 47; DB 2; Length 337;
Best Local Similarity 75.0%; Pred. No. 5.12e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155
| ||:|
QY 57 SAGLHLRV 64

RESULT 7
ENTRY
TITLE      #type complete
#formal_name Pseudomonas aeruginosa
ORGANISM    integrase - Pseudomonas aeruginosa
DATE        #formal_name Pseudomonas aeruginosa
07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
05-Dec-1997
ACCESSIONS  J00300
REFERENCE     J00299
#authors     Wohlleben, W.; Arnold, W.; Bissonnette, L.; Pelletier, A.; Tanquay, A.; Roy, P.H.; Gamboa, G.C.; Barry, G.F.; Aubert, E.; Davies, J.; Kagan, S.A.
#journal     Mol. Gen. Genet. (1989) 217:202-208
#title       On the evolution of Tn21-like multiresistance transposons:
sequence analysis of the gene (aacC1) for gentamicin
acetyltransferase-3-I(AAC(3)-I), another member of the
Tn21-based expression cassette.
#cross-references MUID:89364699
#contents     Plasmid R1033
#accession   J00300
##molecule_type DNA
##residues   1-337 ##label WOH
##cross-references GB:X15852; NID:g45765; PID:g45766; GB:U12338;
NID:g530803; PID:g530804
##note        this sequence is encoded by transposon Tn1696
CLASSIFICATION
#superfamily integrase-like protein
SUMMARY     #length 337 #molecular-weight 38372 #checksum 196

Query Match      83.9%; Score 47; DB 2; Length 337;
Best Local Similarity 75.0%; Pred. No. 5.12e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155
| ||:|
QY 57 SAGLHLRV 64

RESULT 8
ENTRY
TITLE      #type complete
#formal_name Bacillus subtilis
ORGANISM    monooxygenase homolog yhiG - Bacillus subtilis
DATE        #formal_name Bacillus subtilis
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998
ACCESSIONS  F69833
REFERENCE     F69830
#authors     Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Fouiger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinols, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;

```

```

Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogilvara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott, G.;
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Sekowka, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Shln, B.S.; Soldo,
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.

#journal
#title
#cross-references EMBL:98044033
#accession
#status
#molecule_type DNA
#residues 1-499 #label KUN
#cross-references GB:299109; GB:AL009126; NID:g2633260; PID:el183052;
#experimental_source strain 168

GENETICS
#gene
#summary
#length 499 #molecular-weight 54356 #checksum 9354
#type complete
#formal_name Azotobacter vinelandii
#journal
#title
#cross-references EMBL:J03411; NID:g142336; PID:g142340
#length 147 #molecular-weight 16425 #checksum 4849

Query Match 80.4%; Score 45; DB 2; Length 499;
Best Local Similarity 100.0%; Pred. No. 1.40e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 352 AGLHLR 357
QY 58 AGLHLR 63

RESULT 9
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references EMBL:88169465
#accession
#molecule_type DNA
#residues 1-147 #label JOE
#cross-references EMBL:J03411; NID:g142336; PID:g142340
#superfamily drag protein
#length 147 #molecular-weight 16425 #checksum 4849

Query Match 78.6%; Score 44; DB 2; Length 147;
Best Local Similarity 87.5%; Pred. No. 2.29e+01;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 26 SAGLHLV 33
QY 57 SAGLHLV 64

RESULT 10
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references EMBL:86313596
#accession
#molecule_type mRNA
#residues 1-572 #label FIS
#cross-references GB:M13451; NID:g186925; PID:g307108
#Lamins A (see PIR:VEHULA) and C are products of alternative
splicing of the same gene.
The lamins (A, B, and C) contains several alpha-helical domains
capable of forming coiled coils.
The association of lamins, dependent upon ionic interactions, is
interrupted by hyperphosphorylation during the
interphase-metaphase period of the cell cycle. The nuclear
envelope disintegrates with lamin dissociation; it does not
reform until telophase, when the lamins are dephosphorylated to
interphase levels and then reassociate.

GENETICS
#gene
#summary
#length 572 #label FIS
#cross-references GDB:132146; OMIM:150330
#map_position 1q21.2-1q21.3
#description
#structural component of the nuclear lamina, a fibrous
meshwork on the nucleoplasmic surface of the nuclear
membrane

```

```

TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#description
#accession
#status
#molecule_type mRNA
#residues 1-461 #label PON
#cross-references EMBL:248313; NID:g673496; PID:g673497
#length 461 #molecular-weight 51881 #checksum 1885

SUMMARY
Query Match 78.6%; Score 44; DB 2; Length 461;
Best Local Similarity 62.5%; Pred. No. 2.29e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 267 NVGLQLRV 274
QY 57 SAGLHLRV 64

RESULT 11
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references EMBL:86118697
#accession
#molecule_type mRNA
#residues 1-572 #label MCK
#cross-references GB:X03445; NID:g34235; PID:g34236
#A94121
#Fisher, D.Z.; Chaudhary, N.; Blobel, G.
#Proc. Natl. Acad. Sci. U.S.A. (1986) 83:6450-6454
#CDNA sequencing of nuclear lamins A and C reveals primary and
secondary structural homology to intermediate filament
proteins.

#cross-references MUID:86313596
#accession
#molecule_type mRNA
#residues 1-572 #label FIS
#cross-references GB:M13451; NID:g186925; PID:g307108
#Lamins A (see PIR:VEHULA) and C are products of alternative
splicing of the same gene.
The lamins (A, B, and C) contains several alpha-helical domains
capable of forming coiled coils.
The association of lamins, dependent upon ionic interactions, is
interrupted by hyperphosphorylation during the
interphase-metaphase period of the cell cycle. The nuclear
envelope disintegrates with lamin dissociation; it does not
reform until telophase, when the lamins are dephosphorylated to
interphase levels and then reassociate.

GENETICS
#gene
#summary
#length 572 #label FIS
#cross-references GDB:132146; OMIM:150330
#map_position 1q21.2-1q21.3
#description
#structural component of the nuclear lamina, a fibrous
meshwork on the nucleoplasmic surface of the nuclear
membrane

```



```

CLASSIFICATION #superfamily cytoskeletal keratin
KEYWORDS alternative splicing; coiled coil; heptad repeat; membrane
protein; nuclear membrane; phosphoprotein; structural
protein

FEATURE
2-572 #product lamin C #status predicted #label MAT\
2-33 #domain head #label ROD\
34-388 #domain rod #label ROD\
34-70 #region coil 1A\
81-218 #region coil 1B\
243-388 #region coil 2\
266 #region heptad change of phase\
330 #region heptad change of phase\
417-420 #domain tail #label END\
SUMMARY #length 572 #molecular-weight 65134 #checksum 6769

Query Match 78.6%; Score 44; DB 1; Length 572;
Best Local Similarity 62.5%; Pred. No. 2.29e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 56 NAGLRLRI 63
:||||:
Qy 57 SAGLHLRV 64

RESULT 12
ENTRY S04333 #type complete
TITLE lamin C - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
17-Mar-1999
ACCESSIONS S04333; S03704; S65931
REFERENCE Riedel, W.; Werner, D.
#authors Biochim. Biophys. Acta (1989) 1008:119-122
#journal Nucleotide sequence of the full-length mouse lamin C cDNA and
#title its deduced amino-acid sequence.
#cross-references MUID:89247440
#accession S04333
##molecule_type mRNA
##residues 1-574 ##label RTE
##cross-references EMBL:X14170; NID:g52864; PID:g52865
REFERENCE S06662
#authors Weber, K.; Plessmann, U.; Traub, P.
#journal FEBS Lett. (1989) 257:411-414
#title Maturation of nuclear lamin A involves a specific
carboxy-terminal trimming, which removes the
polyisoprenylation site from the precursor; implications
for the structure of the nuclear lamina.
#cross-references MUID:90060368
#accession S09704
##molecule_type protein
##residues 521-574 ##label WEB
REFERENCE S65931
#authors Nakajima, N.; Abe, K.
#journal FEBS Lett. (1995) 365:108-114
#title Genomic structure of the mouse A-type lamin gene locus
encoding somatic and germ cell-specific lamins.
#cross-references MUID:95300954
#accession S65931
##status translation not shown
##molecule_type DNA
##residues 1-117 ##label NAK
##cross-references EMBL:D49732; NID:g1100724
CLASSIFICATION #superfamily cytoskeletal keratin
KEYWORDS alternative splicing; coiled coil; heptad repeat;
intermediate filament; nucleus; phosphoprotein

FEATURE
2-33 #domain head #status predicted #label HED\
34-388 #domain rod #status predicted #label ROD\
34-218 #region coiled coil 1\
243-388 #region coiled coil 2\

```

```

266 #region heptad change of phase\
325 #region stutter\
330 #region heptad change of phase\
389-574 #domain tail #status predicted #label END\
417-420 #region nuclear location signal
SUMMARY #length 574 #molecular-weight 65405 #checksum 7573

Query Match 78.6%; Score 44; DB 2; Length 574;
Best Local Similarity 62.5%; Pred. No. 2.29e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 56 NAGLRLRI 63
:||||:
Qy 57 SAGLHLRV 64

RESULT 13
ENTRY S05517 #type complete
TITLE lamin - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
08-Sep-1997
ACCESSIONS S05517
REFERENCE Peter, M.; Kitten, G.T.; Lehner, C.F.; Vorbürger, K.; Bailler,
S.M.; Maridor, G.; Nigg, E.A.
#authors J. Mol. Biol. (1989) 208:393-404
#journal Cloning and sequencing of cDNA clones encoding chicken lamins
#title A and B(1) and comparison of the primary structures of
vertebrate A- and B-type lamins.
#cross-references MUID:90012208
#accession S05517
##status preliminary
##molecule_type mRNA
##residues 1-657 ##label PET
##cross-references EMBL:X16879; NID:g63555; PID:g63556
CLASSIFICATION #superfamily cytoskeletal keratin
KEYWORDS nucleus
SUMMARY #length 657 #molecular-weight 73164 #checksum 1468

Query Match 78.6%; Score 44; DB 2; Length 657;
Best Local Similarity 62.5%; Pred. No. 2.29e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 55 NAGLRLRI 62
:||||:
Qy 57 SAGLHLRV 64

RESULT 14
ENTRY VEHULA #type complete
TITLE lamin A - human
ALTERNATE_NAMES 70KDa lamin
ORGANISM #formal_name Homo sapiens #common_name man
DATE 28-May-1986 #sequence_revision 04-Dec-1986 #text_change
05-Dec-1998
ACCESSIONS A02961; B24249; C24249
REFERENCE A02962
#authors Mkeon, F.D.; Kirschner, M.W.; Caput, D.
#journal Nature (1986) 319:463-468
#title Homologies in both primary and secondary structure between
nuclear envelope and intermediate filament proteins.
#cross-references MUID:86118697
#accession A02961
##molecule_type mRNA
##residues 1-582, 'LAHRAVRLRAACRQICQRLRSPGGRTHL',
'LWLFCLQCHGHSQLPQCGGQWQLRGQSG',
'HPLPPGLOLPPNPPELQHHVLDLPGR',
'GGGGFLRPPLMPPCPARHGRLGAEK' ##label MCK
##cross-references GB:X03444; NID:g34227; PID:g34228
##note this sequence has been corrected in reference A94121; an
omitted nucleotide caused a reading frame error
beginning at position 583

```

```

REFERENCE
#authors Fisher, D.Z.; Chaudhary, N.; Blobel, G.
#journal Proc. Natl. Acad. Sci. U.S.A. (1986) 83:6450-6454
#title cDNA sequencing of nuclear lamins A and C reveals primary and
        secondary structural homology to intermediate filament
        proteins.
#cross-references MUID:86313596
#note parts of sequences from rat lamins, but not human, were
        determined by protein sequencing
#accession B24249
##molecule_type mRNA
##residues 537-664 #label FIS
##note sequence fragment shown in publication
#accession C24249
##molecule_type mRNA
##residues 150-664 #label FI2
##cross-references GB:M13452; NID:9186838; PID:9386856
##note submitted sequence extracted from GenBank
#COMMENT Lamins A and C (see PIR:VEHULC) are products of alternative
        splicing of the same gene.
#COMMENT The lamins (A, B, and C) contains several alpha-helical domains
        capable of forming coiled coils.
#COMMENT The association of lamins, dependent upon ionic interactions, is
        interrupted by hyperphosphorylation during the
        interphase-metaphase period of the cell cycle. The nuclear
        envelope disintegrates with lamin dissociation; it does not
        reform until telophase, when the lamins are dephosphorylated to
        intermediate levels and then reassociate.

GENETICS
#gene GDB:LMNA; LMN1
##cross-references GDB:I32146; OMIM:150330
#map_position lq21.2-lq21.3
#FUNCTION
#description structural component of the nuclear lamina, a fibrous
        meshwork on the nucleoplasmic surface of the nuclear
        membrane
#CLASSIFICATION
#superfamily cytoskeletal keratin
#KEYWORDS
        alternative splicing; coiled coil; heptad repeat;
        lipoprotein; membrane protein; methylated carboxyl end;
        nuclear membrane; phosphoprotein; prenylated cysteine;
        structural protein
#FEATURE
2-661 #product lamin A #status predicted #label MAT\
2-33 #domain head #label HED\
34-388 #domain rod #label ROD\
34-70 #region coil 1A\
81-218 #region coil 1B\
243-388 #region coil 2\
325 #region stutter\
389-661 #domain tail #label END\
417-420 #region nuclear location signal\
661 #binding_site farnesyl (Cys) (covalent) #status
        predicted\
661 #modified_site methyl ester carboxyl end (Cys) (in
        mature form) #status predicted
#SUMMARY #length 664 #molecular-weight 74139 #checksum 9955
Query Match 78.6%; Score 44; DB 1; Length 664;
Best Local Similarity 62.5%; Pred. No. 2-29e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 56 NAGLRLRI 63
QY 57 SAGLHLRV 64
:||||:

RESULT 15
ENTRY #type complete
TITLE lamin A - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
17-Mar-1999
ACCESSIONS S28182; S06662; S65931

```

```

REFERENCE
#authors Nakajima, N.; Sado, T.
#journal Blochim. Biophys. Acta (1993) 1171:311-314
#title Nucleotide sequence of a mouse lamin A cDNA and its deduced
        amino acid sequence.
#cross-references MUID:93144345
#accession S28182
##status preliminary
##molecule_type mRNA
##residues 1-665 #label NAK
#REFERENCE S06662
#authors Weber, K.; Plessmann, U.; Traub, P.
#journal FEBS Lett. (1989) 257:411-414
#title Maturation of nuclear lamin A involves a specific
        carboxy-terminal trimming, which removes the
        polyisoprenylation site from the precursor; implications
        for the structure of the nuclear lamina.
#cross-references MUID:90060368
#accession S06662
##molecule_type protein
##residues 520-580, 'X', 582-622, 'V', 624-647 #label WEB
#REFERENCE S65931
#authors Nakajima, N.; Abe, K.
#journal FEBS Lett. (1995) 365:108-114
#title Genomic structure of the mouse A-type lamin gene locus
        encoding somatic and germ cell-specific lamins.
#cross-references MUID:95300954
#accession S65931
##status translation not shown
##molecule_type DNA
##residues 1-117 #label NAW
##cross-references EMBL:D49732; NID:g1100724
#CLASSIFICATION
#superfamily cytoskeletal keratin
#KEYWORDS
        alternative splicing; coiled coil; heptad repeat;
        intermediate filament; lipoprotein; nucleus;
        phosphoprotein; prenylated cysteine
#FEATURE
1-33 #domain head #status predicted #label HED\
34-388 #domain rod #status predicted #label ROD\
34-218 #region coiled coil 1\
243-388 #region coiled coil 2\
266 #region heptad change of phase\
325 #region stutter\
330 #region heptad change of phase\
389-662 #domain tail #status predicted #label END\
417-420 #region nuclear location signal\
662 #binding_site farnesyl (Cys) (covalent) #status
        predicted
#SUMMARY #length 665 #molecular-weight 74184 #checksum 676
Query Match 78.6%; Score 44; DB 2; Length 665;
Best Local Similarity 62.5%; Pred. No. 2-29e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 56 NAGLRLRI 63
QY 57 SAGLHLRV 64
:||||:

Search completed: Wed Sep 1 16:27:57 1999
Job time : 12 secs.

```

\*\*\*\*\*  
WIREH  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Sep 1 16:26:37 1999; MasPar time 2.52 Seconds  
Tabular output not generated. 89.698 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (57-64) from PCTUS9913024.pep (10 of 12)  
Perfect Score: 56  
Sequence: 1 SAGLHLRV 8

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 22.517; Variance 22.845; scale 0.986

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	ID	Description	Pred. No.
1	47	83.9	1	INTR_ECOLI	2.31e+00
2	44	78.6	1	RBL_PTEVI	1.16e+01
3	44	78.6	1	DVR1_STRPU	1.16e+01
4	44	78.6	1	LAMC_HUMAN	1.16e+01
5	44	78.6	1	LAMC_MOUSE	1.16e+01
6	44	78.6	1	LAMA_CHICK	1.16e+01
7	44	78.6	1	LAMA_HUMAN	1.16e+01
8	44	78.6	1	LAMA_MOUSE	1.16e+01
9	44	78.6	1	LAMA_RAT	1.16e+01
10	43	76.8	1	PHALATE 4,5-DIOXYGEN	1.94e+01
11	43	76.8	1	PH2_PSEPU	1.94e+01
12	43	76.8	1	RP55_RHIME	1.94e+01
13	43	76.8	1	RP54_RHIME	1.94e+01
14	43	76.8	1	RP54_RHISN	1.94e+01
15	43	76.8	1	DESMOGLIN 2 PRECURSOR	1.94e+01
16	42	75.0	1	VIT6_OSCBR	1.94e+01
17	42	75.0	1	HS2M_SOYBN	3.23e+01
18	42	75.0	1	NECD_MOUSE	3.23e+01
19	42	75.0	1	MEP3_YEAST	3.23e+01
20	42	75.0	1	ACH2_RAT	3.23e+01
21	42	75.0	1	NEURONAL ACETYLCHOLINE	3.23e+01
22	42	75.0	1	CYTCHROME P450 4E5 PR	3.23e+01
23	42	75.0	1	FTHW_MYCTU	3.23e+01
24	42	75.0	1	HYPOTHETICAL 56.3 KD F	3.23e+01
25	42	75.0	1	HCVD_EURCA	3.23e+01
26	42	75.0	1	HEMOCYANIN D CHAIN	3.23e+01
27	42	75.0	1	PROTEIN-GLUTAMINE GAMM	3.23e+01

24 42 75.0 702 1 YCBY\_ECOLI HYPOTHETICAL 78.9 KD P 3.23e+01  
25 42 75.0 862 1 MUTS\_BORBU DNA MISMATCH REPAIR PR 3.23e+01  
26 42 75.0 1072 1 HSER\_RAT HEAT-STABLE ENTEROTOXI 3.23e+01  
27 42 75.0 1073 1 HSER\_PIG HEAT-STABLE ENTEROTOXI 3.23e+01  
28 42 75.0 1073 1 HSER\_HUMAN HEAT-STABLE ENTEROTOXI 3.23e+01  
29 42 75.0 1076 1 HSER\_CAVPO HEAT-STABLE ENTEROTOXI 3.23e+01  
30 41 73.2 72 1 YCA1\_OENBE HYPOTHETICAL 8.5 KD PR 5.31e+01  
31 41 73.2 149 1 Y213\_METJA HYPOTHETICAL PROTEIN M 5.31e+01  
32 41 73.2 214 1 RL10\_CAEEL 60S RIBOSOMAL PROTEIN 5.31e+01  
33 41 73.2 228 1 RL10\_PINTA 60S RIBOSOMAL PROTEIN 5.31e+01  
34 41 73.2 247 1 ZO20\_XENLA OOCYTE ZINC FINGER PRO 5.31e+01  
35 41 73.2 328 1 Y052\_HAEIN HYPOTHETICAL PROTEIN H 5.31e+01  
36 41 73.2 340 1 Y0BN\_MYCTU HYPOTHETICAL 37.2 KD P 5.31e+01  
37 41 73.2 364 1 F16P\_XANFL FRUCTOSE-1,6-BISPHOSPH 5.31e+01  
38 41 73.2 426 1 SLS1\_YARLI SLS1 PROTEIN PRECURSOR 5.31e+01  
39 41 73.2 433 1 DHOM\_SYNY3 HOMOSERINE DEHYDROGENA 5.31e+01  
40 41 73.2 563 1 METB\_ARATH CYSTATHIONINE GAMMA-SY 5.31e+01  
41 41 73.2 611 1 YX45\_MYCTU HYPOTHETICAL 67.2 KD P 5.31e+01  
42 41 73.2 1130 1 YL17\_CAEEL HYPOTHETICAL 131.5 KD 5.31e+01  
43 41 73.2 1308 1 YTX2\_XENLA TRANSPONSON TX1 HYPOTHE 5.31e+01  
44 41 73.2 1547 1 RRPO\_PMV RNA REPLICATION PROTEI 5.31e+01  
45 41 73.2 1841 1 SUI5\_RAT SUCRASE-ISOMALTASE, IN 5.31e+01

## ALIGNMENTS

RESULT 1  
ID INTR\_ECOLI STANDARD; PRT; 337 AA.  
AC P09999; P11743;  
DT 01-MAR-1989 (REL. 10, CREATED)  
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE INTEGRASE/RECOMBINASE (E2 PROTEIN).  
GN INT.  
OS ESCHERICHIA COLI, AND PSEUDOMONAS AERUGINOSA.  
OG PLASMID INCN R46, PLASMID RGN238, PLASMID RI033, PLASMID PLMO20,  
OG PLASMID PLMO27, PLASMID PLMO150, PLASMID PLMO229, AND PLASMID PVS1.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
OC ESCHERICHIA.  
RN [1]  
RC SEQUENCE FROM N.A.  
RX PLASMID-INCN R46;  
RX MEDLINE; 88015610.  
RA HALL R.M., VOCKLER C.;  
RT "The region of the IncN plasmid R46 coding for resistance to  
beta-lactam antibiotics, streptomycin/spectinomycin and sulphonamides  
is closely related to antibiotic resistance segments found in IncW  
plasmids and in Tn21-like transposons.";  
RT NUCLEIC ACIDS RES. 15:7491-7501(1987).  
RN [2]  
RP PARTIAL SEQUENCE FROM N.A.  
RC PLASMID-RGN238;  
RX MEDLINE; 88041110.  
RA OUELLETTE M., BISSONNETTE L., ROY P.H.;  
RT "Precise insertion of antibiotic resistance determinants into  
Tn21-like transposons: nucleotide sequence of the OXA-1  
beta-lactamase gene.";  
RT PROC. NATL. ACAD. SCI. U.S.A. 84:7378-7382(1987).  
RN [3]  
RC SEQUENCE FROM N.A.  
RX PLASMID-RGN238;  
RA OUELLETTE M.;  
RT SUBMITTED (DEC-1987) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [4]  
RC SEQUENCE FROM N.A.  
RX PLASMID-PLMO20;  
RA MEDLINE; 89039710.  
RT SUNDTSTROM L., RADSTROM P., SWEDBERG G., SKOELD O.;  
RT "Site-specific recombination promotes linkage between trimethoprim-  
and sulfonamide resistance genes. Sequence characterization of dhfrv  
and sulI and a recombination active locus of Tn21.";  
RL MOL. GEN. GENET. 213:191-201(1988).  
RN [5]

SEQUENCE FROM N.A.  
RC SPECIES-P.AERUGINOSA; PLASMID-R1033; TRANSPONSON-TN1696;  
RX MEDLINE; 89364699.  
RA WOHLLEBEN W., ARNOLD W., BISSONNETTE L., PALLETIER A., TANGUAY A.,  
ROY P.H., GAMBOA G.C., BARRY G.F., AUBERT E., DAVIES J., KAGAN S.A.;  
RT "On the evolution of *tn21*-like multiresistance transposons: sequence  
analysis of the gene (aacC1) for gentamicin acetyltransferase-3-  
1(AAC(3)-1), another member of the *Tn21*-based expression cassette.";  
RL MOL. GEN. GENET. 217:202-208(1989).  
RN [6]  
RP SEQUENCE OF 1-145 FROM N.A.  
RC PLASMID-PLM0150, AND PLM0229;  
RX MEDLINE; 90262183.  
RA SUNDBLOM L., SKOELD O.;  
RT "The *dhfr1* trimethoprim resistance gene of *Tn7* can be found at  
specific sites in other genetic surroundings.";  
RL ANTIMICROB. AGENTS CHEMOTHER. 34:642-650(1990).  
RN [7]  
RP SEQUENCE FROM N.A.  
RC SPECIES-P.AERUGINOSA; PLASMID-PVSL1;  
RX MEDLINE; 92138617.  
RA BISSONNETTE L., ROY P.H.;  
RT "Characterization of *In0* of *Pseudomonas aeruginosa* plasmid pVSL1, an  
ancestor of integrons of multiresistance plasmids and transposons of  
gram-negative bacteria.";  
RL J. BACTERIOL. 174:1248-1257(1992).  
RN [8]  
RP SEQUENCE OF 1-145 FROM N.A.  
RC SPECIES-E.COLI; PLASMID-PLM027; TRANSPONSON-TN5086;  
RA SUNDBLOM L., SWEDBERG G., SKOELD O.;  
RL SUBMITTED (MAR-1991) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- FUNCTION: PUTATIVE INTEGRASE BELIEVED TO BE INVOLVED IN  
INSERTIONS OF ANTIBIOTIC RESISTANCE GENES INTO PLASMIDS  
AND TRANSPONSONS.  
CC -!- THE SEQUENCE SHOWN IS THAT OF PLASMID R46.  
CC -!- SIMILARITY: BELONGS TO THE "PHAGE" INTEGRASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U12338; G530804;  
DR EMBL; U49101; G530814;  
DR EMBL; X06046; G42207;  
DR EMBL; M95287; G151817;  
DR EMBL; J02957; G152063;  
DR EMBL; X12868; G45673;  
DR EMBL; X15852; G45766;  
DR EMBL; X12870; G482067;  
DR EMBL; X17478; G45584;  
DR EMBL; X17477; G45579;  
DR EMBL; M73819; G151300;  
DR EMBL; X58425; G43091;  
DR PIR; C26839; C26839.  
DR PIR; A42646; A42646.  
DR PFAM; PF00589; Phage\_integrase; 1.  
DR HSSP; P21891; LAOP.  
KW DNA RECOMBINATION; DNA INTEGRATION; PLASMID; TRANSPONSONABLE ELEMENT.  
FT VARIANT 32 32 R -> P (IN PLM0150).  
FT VARIANT 39 39 H -> N (IN R1033).  
FT VARIANT 82 82 A -> P (IN PLM020).  
FT VARIANT 84 84 A -> P (IN PLM020).  
FT VARIANT 187 187 S -> T (IN R1033).  
FT VARIANT 205 205 E -> Q (IN RGN238).  
FT ACT\_SITE 312 312 PROBABLE TRANSIENT COVALENT LINKAGE TO  
DNA DURING STRAND CLEAVAGE AND REJOINING.  
SQ SEQUENCE 337 AA; 38381 MW; C7BE628D CRC32;  
Query Match 83.9%; Score 47; DB 1; Length 337;  
Best Local Similarity 75.0%; Pred. No. 2.31e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Db 148 SEGQLRV 155  
QY 57 SAGHLRV 64  
RESULT 2  
ID RBL\_PTEVI STANDARD; PRT; 440 AA.  
AC Q33015;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO  
LARGE SUBUNIT) (FRAGMENT).  
GN RBCL.  
OS PTERIS VITTATA (CHINESE BRAKE).  
OG CHLOROPLAST.  
OC EUPHYLOTIA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
OC EUPHYLOPHYTES; FILICOPHYTA; FILICOPSIDA; FILICALES; PTERIDACEAE;  
OC PTERIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LEAF;  
RX MEDLINE; 95211278.  
RA WOLF P.G., SOLTIS P.S., SOLTIS D.E.;  
RT "Phylogenetic relationships of *demstaedtioid* ferns: evidence from  
*rbcl* sequences";  
RL MOL. PHYLOGENET. EVOL. 3:383-392(1994).  
CC -!- FUNCTION: RUBISCO CATALYSES TWO REACTIONS: THE CARBOXYLATION OF  
CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
ACTIVE SITE.  
CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) -  
2 3-PHOSPHO-D-GLYCERATE.  
CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) -  
3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCULATE.  
CC -!- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.  
CC -!- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U05941; G463542;  
DR PROSITE; PS00157; RUBISCO\_LARGE; 1.  
DR PFAM; PF00016; RUBISCO\_large; 1.  
DR MENDEL; 3802; PTEVI:rbcl;1.  
DR HSSP; P00875; LAUS.  
KW PHOTOSYNTHESIS; CARBON DIOXIDE FIXATION; PHOTORESPIRATION;  
LYASE; OXIDOREDUCTASE; MONOOXYGENASE; CHLOROPLAST.  
FT NON\_TER 1 1  
FT ACT\_SITE 191 191  
FT NON\_TER 440 440  
SQ SEQUENCE 440 AA; 48572 MW; BDE52A9 CRC32;  
Query Match 78.6%; Score 44; DB 1; Length 440;  
Best Local Similarity 83.3%; Pred. No. 1.16e+01;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Db 298 GMLRV 303  
QY 59 GHLRV 64

```
RESULT 3
ID DVL1_STRPU STANDARD; PRT; 461 AA.
AC P4899;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE DVL1 PROTEIN HOMOLOG PRECURSOR.
GN DVL1
OS STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC ECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC STRONGYLOCENTROTUS.
RN [1]
RP SEQUENCE FROM N.A.
RA PONCE M.R., MICOL J.L., DAVIDSON E.H.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (PROBABLY).
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z48313; G673497;
CC PROSITE; PS00250; TGF-BETA; 1.
CC PFAM; PF00019; TGF-beta; 1.
CC HSP; P18075; IBMP.
CC GROWTH FACTOR; CYTOKINE; GLYCOPROTEIN; SIGNAL.
CC SIGNAL 1 30
CC PROPEP 31 338
CC CHAIN 339 461
CC DISULFID 360 426
CC DISULFID 389 458
CC DISULFID 393 460
CC DISULFID 425 425
CC CARBOHYD 149 149
CC CARBOHYD 402 402
CC SEQUENCE 461 AA; 51881 MW; 57C7CBA0 CRC32;
CC -----
Query Match 78.6%; Score 44; DB 1; Length 461;
Best Local Similarity 62.5%; Pred. No. 1.16e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 267 NVGLQLEV 274
QY 57 SAGLHLRV 64
RESULT 4
ID LAMC_HUMAN STANDARD; PRT; 572 AA.
AC P02546;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE LAMIN C.
GN LAMIN C.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA MCKEON F.D., KIRSCHNER M.W., CAPUT D.;
RL "Homologies in both primary and secondary structure between nuclear
RL envelope and intermediate filament proteins."
RL NATURE 319:463-468(1986).
RN [2]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE; 86313596.
RA FISHER D.Z., CHAUDHARY N., BLOBEL G.;
RT "cDNA sequencing of nuclear lamins A and C reveals primary and
RT secondary structural homology to intermediate filament proteins.";
RL PROC. NATL. ACAD. SCI. U.S.A. 83:6450-6454(1986).
CC -1- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS
CC LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,
CC WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE
CC AND MAY ALSO INTERACT WITH CHROMATIN.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- PTM: INCREASED PHOSPHORYLATION OF THE LAMINS OCCURS BEFORE
CC ENVELOPE DISINTEGRATION AND PROBABLY PLAYS A ROLE IN REGULATING
CC LAMIN ASSOCIATIONS.
CC -1- THERE ARE THREE TYPES OF LAMINS IN HUMAN CELLS: A, B, AND C.
CC -1- THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY CONTROLLED BY
CC THE CELL CYCLE, AS SEEN BY THE DISINTEGRATION AND FORMATION OF
CC THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE, RESPECTIVELY.
CC -1- ALTERNATIVE PRODUCTS: LAMINS A AND C ARE THE PRODUCTS OF ALTERNATE
CC SPLICING OF THE SAME GENE. THEY ARE PRESENT IN EQUAL AMOUNTS IN
CC THE LAMINA OF MAMMALS.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M13451; G307108;
CC EMBL; X03445; G34236;
CC PIR; A02962; VEHULC.
CC PIR; A24249; A24249.
CC MW; 150330;
CC PROSITE; PS00226; IF; 1.
CC PFAM; PF00038; filament; 1.
CC PFAM; PF00932; IF-tail; 1.
CC INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL;
CC NUCLEAR PROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING.
CC DOMAIN 1 33
CC DOMAIN 34 383
CC DOMAIN 384 572
CC ROD.
CC COIL 1A.
CC COIL 1B.
CC LINKER 1.
CC COIL 1B.
CC LINKER 2.
CC COIL 2.
CC STUTTER (BY SIMILARITY).
CC HEPTAD CHANGE OF PHASE.
CC HEPTAD CHANGE OF PHASE.
CC NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC SEQUENCE 572 AA; 65134 MW; 6F6A20D0 CRC32;
CC -----
Query Match 78.6%; Score 44; DB 1; Length 572;
Best Local Similarity 62.5%; Pred. No. 1.16e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 56 NAGLRLRI 63
QY 57 SAGLHLRV 64
RESULT 5
ID LAMC_MOUSE STANDARD; PRT; 574 AA.
AC P11516;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE LAMINS C AND C2.
GN LMN1 OR LMNA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
```

CC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89247440.  
 RA RIEDEL W., WERNER D.;  
 RT "Nucleotide sequence of the full-length mouse lamin C cDNA and its  
 RL deduced amino-acid sequence.";  
 RL BIOCHIM. BIOPHYS. ACTA 1008:119-122(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A. (LAMIN C2).  
 RC STRAIN-DDY; TISSUE-TESTIS;  
 RX MEDLINE; 94244715  
 RA FURUKAWA K., INAGAKI H., HOTTA Y.;  
 RT "Identification and cloning of an mRNA coding for a germ  
 RL cell-specific A-type lamin in mice.";  
 RL EXP. CELL RES. 212:426-430(1994).  
 RN [3]  
 RP SEQUENCE OF 235-568 FROM N.A.  
 RX MEDLINE; 93144345.  
 RA NAKAJIMA N., SADO T.;  
 RT "Nucleotide sequence of a mouse lamin A cDNA and its deduced amino  
 RL acid sequence.";  
 RL BIOCHIM. BIOPHYS. ACTA 1171:311-314(1993).  
 RN [4]  
 RP SEQUENCE OF 1-119 FROM N.A.  
 RX MEDLINE; 95300954.  
 RA NAKAJIMA N., ABE K.;  
 RT "Genomic structure of the mouse A-type lamin gene locus encoding  
 RL somatic and germ cell-specific lamins.";  
 RL FEBS LETT. 365:108-114(1995).  
 RN [5]  
 RP PARTIAL SEQUENCE, AND PHOSPHORYLATION.  
 RX MEDLINE; 92070490.  
 RA EGGERT M., RADOMSKI N., TRIPIER D., TRAUB P., JOST E.;  
 RT "Identification of phosphorylation sites on murine nuclear lamin C by  
 RL RP-HPLC and microsequencing.";  
 RL FEBS LETT. 292:205-209(1991).  
 CC -1- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS  
 CC LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,  
 CC WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE  
 CC AND MAY ALSO INTERACT WITH CHROMATIN.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: LAMIN C2 IS SPECIFICALLY EXPRESSED IN GERM  
 CC CELLS. THIS SPECIFIC EXPRESSION AND UNIQUE STRUCTURE SUGGESTS A  
 CC ROLE FOR LAMIN C2 IN DETERMINING THE ORGANIZATION OF NUCLEAR AND  
 CC CHROMOSOMAL STRUCTURES DURING SPERMATOGENESIS.  
 CC -1- PTM: INCREASED PHOSPHORYLATION OF THE LAMINS OCCURS BEFORE  
 CC ENVELOPE DISINTEGRATION AND PROBABLY PLAYS A ROLE IN REGULATING  
 CC LAMIN ASSOCIATIONS.  
 CC -1- THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY CONTROLLED BY  
 CC THE CELL CYCLE, AS SEEN BY THE DISINTEGRATION AND FORMATION OF  
 CC THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE, RESPECTIVELY.  
 CC -1- ALTERNATIVE PRODUCTS: LAMINS A, C AND C2 ARE THE PRODUCTS OF  
 CC ALTERNATE SPLICING OF THE SAME GENE. LAMINS A AND C ARE PRESENT IN  
 CC EQUAL AMOUNTS IN THE LAMINA OF MAMMALS.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X14170; G52865;  
 CC EMBL; D14850; G505027;  
 CC EMBL; D49732; E212319;  
 CC EMBL; D49733; G1794160;  
 CC EMBL; D49733; G1794161;  
 CC EMBL; S04333; S04333;  
 CC PIR; S18324; S18324;  
 CC MGD; MGI:96794; LMNA.

DR PROSITE; PS00226; IF; 1.  
 DR PFAM; PF00038; filament; 1.  
 DR PFAM; PF00932; IF\_tail; 1.  
 KW INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL;  
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING.  
 FT DOMAIN 1 33 HEAD.  
 FT DOMAIN 34 383 ROD.  
 FT DOMAIN 384 574 TAIL.  
 FT DOMAIN 34 70 COIL 1A.  
 FT DOMAIN 71 80 LINKER 1.  
 FT DOMAIN 81 218 COIL 1B.  
 FT DOMAIN 219 242 LINKER 2.  
 FT DOMAIN 243 383 COIL 2.  
 FT SITE 325 325 STUTTER (BY SIMILARITY).  
 FT SITE 266 266 HEPTAD CHANGE OF PHASE.  
 FT SITE 330 330 HEPTAD CHANGE OF PHASE.  
 FT SITE 401 401 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 417 422 PHOSPHORYLATION (BY CDC2) (PARTIAL,  
 FT MOD\_RES 392 392 AT INTERPHASE).  
 FT MOD\_RES 407 407 PHOSPHORYLATION (PARTIAL, AT INTERPHASE).  
 FT MOD\_RES 409 409 PHOSPHORYLATION (PARTIAL, AT INTERPHASE).  
 FT VARSPLIC 1 112 MISSING (IN LAMIN C2).  
 FT VARSPLIC 113 118 FKELKA -> MGNREG (IN LAMIN C2).  
 FT CONFLICT 118 119 AR -> VC (IN REF. 1).  
 FT CONFLICT 401 401 R -> P (IN REF. 1).  
 FT CONFLICT 439 440 RV -> WL (IN REF. 1).  
 SQ SEQUENCE 574 AA; 65446 MW; 3C8AD85A CRC32;  
 Query Match 78.6%; Score 44; DB 1; Length 574;  
 Best Local Similarity 62.5%; Pred. No. 1.16e+01;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 Db 56 NAGLRLRI 63  
 Qy 57 SAGLRLRV 64  
 RESULT 6  
 ID LAMA\_CHICK STANDARD; PRT; 657 AA.  
 AC P13648;  
 DT 01-JAN-1990 (REL. 13, CREATED)  
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE LAMIN A.  
 OS GALLUS GALLUS (CHICKEN).  
 OC EUKARYOTA; METAZOA; VERTEBRATA; ARCHOSAURIA; AVES;  
 OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 90012208.  
 RA PETER M., KITTEL G.T., LEHNER C.F., VORBURGER K., BAILER S.M.,  
 RA MARIDOR G., NIGG E.A.;  
 RT "Cloning and sequencing of cDNA clones encoding chicken lamins A and  
 RL B1 and comparison of the primary structures of vertebrate A- and  
 RL B-type lamins.";  
 RL J. MOL. BIOL. 208:393-404(1989).  
 CC -1- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS  
 CC LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,  
 CC WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE  
 CC AND MAY ALSO INTERACT WITH CHROMATIN.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X16879; G63556;  
 CC PIR; S05517; S05517.

DR PROSITE; PS00226; IF; 1.  
DR PFAM; PF00038; filament; 1.  
DR PFAM; PF00932; IF tail; 1.  
KW INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL;  
KW NUCLEAR PROTEIN; LIPOPROTEIN; PRENYLATION; PHOSPHORYLATION;  
KW ALTERNATIVE SPLICING.  
FT DOMAIN 1 32 HEAD.  
FT DOMAIN 33 382 ROD.  
FT DOMAIN 383 657 TAIL.  
FT DOMAIN 33 69 COIL 1A.  
FT DOMAIN 70 79 LINKER 1.  
FT DOMAIN 80 217 LINKER 2.  
FT DOMAIN 218 241 LINKER 2.  
FT DOMAIN 242 382 COIL 2.  
FT LIPID 654 654 FARNESYL (BY SIMILARITY).  
SQ SEQUENCE 657 AA; 73164 MW; 3E7C5858 CRC32;  
  
Query Match 78.6%; Score 44; DB 1; Length 657;  
Best Local Similarity 62.5%; Pred. No. 1.16e+01;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
Db 55 NAGLRLRI 62  
QY 57 SAGLHLRV 64  
:|||||:  
57 SAGLHLRV 64  
  
RESULT 7  
ID LAMA HUMAN STANDARD; PRT; 664 AA.  
AC P02345;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 20-MAR-1987 (REL. 04, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE LAMIN A (70 KD LAMIN).  
GN LMN1 OR LMNA.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 86118697.  
RA MCKEON F.D., KIRSCHNER M.W., CAPUT D.;  
RT "Homologies in both primary and secondary structure between nuclear envelope and intermediate filament proteins.";  
RL NATURE 319:463-468(1986).  
RN [2]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 583-644.  
RX MEDLINE; 86313596.  
RA FISHER D.Z., CHAUDHARY N., BLOBEL G.;  
RT "CDNA sequencing of nuclear lamins A and C reveals primary and secondary structural homology to intermediate filament proteins.";  
RL PROC. NATL. ACAD. SCI. U.S.A. 83:6450-6454(1986).  
CC -!- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE, WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE AND MAY ALSO INTERACT WITH CHROMATIN.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- PTM: INCREASED PHOSPHORYLATION OF THE LAMINS OCCURS BEFORE ENVELOPE DISINTEGRATION AND PROBABLY PLAYS A ROLE IN REGULATING LAMIN ASSOCIATIONS.  
CC -!- THERE ARE THREE TYPES OF LAMINS IN HUMAN CELLS: A, B, AND C.  
CC -!- THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY CONTROLLED BY THE CELL CYCLE, AS SEEN BY THE DISINTEGRATION AND FORMATION OF THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE, RESPECTIVELY.  
CC -!- ALTERNATIVE PRODUCTS: LAMINS A AND C ARE THE PRODUCTS OF ALTERNATE SPLICING OF THE SAME GENE. THEY ARE PRESENT IN EQUAL AMOUNTS IN THE LAMINA OF MAMMALS.  
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 582 ONWARD AND IS LONGER (702 AA) DUE TO A FRAMESHIFT.  
  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X03444; G34228; ALT\_FRAME.  
DR EMBL; M13452; G386856;  
DR PIR; A02961; VERULA.  
DR PIR; B24249; B24249.  
DR MIM; I50330;  
DR PROSITE; PS00226; IF; 1.  
DR PFAM; PF00038; filament; 1.  
DR PFAM; PF00932; IF tail; 1.  
KW INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL;  
KW NUCLEAR PROTEIN; LIPOPROTEIN; PRENYLATION; PHOSPHORYLATION;  
KW ALTERNATIVE SPLICING.  
FT MOD\_RES 1 1 BLOCKED.  
FT DOMAIN 1 33 HEAD.  
FT DOMAIN 34 383 ROD.  
FT DOMAIN 384 664 TAIL.  
FT DOMAIN 34 70 COIL 1A.  
FT DOMAIN 81 218 LINKER 1.  
FT DOMAIN 219 242 COIL 1B.  
FT DOMAIN 243 383 LINKER 2.  
FT SITE 325 325 STUTTER (BY SIMILARITY).  
FT DOMAIN 417 422 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT LIPID 661 661 FARNESYL (BY SIMILARITY).  
SQ SEQUENCE 664 AA; 74139 MW; 31F253EE CRC32;  
  
Query Match 78.6%; Score 44; DB 1; Length 664;  
Best Local Similarity 62.5%; Pred. No. 1.16e+01;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
Db 56 NAGLRLRI 63  
QY 57 SAGLHLRV 64  
:|||||:  
57 SAGLHLRV 64  
  
RESULT 8  
ID LAMA\_MOUSE STANDARD; PRT; 665 AA.  
AC P48678; P97859;  
DT 01-FEB-1996 (REL. 33, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE LAMIN A.  
GN LMN1 OR LMNA.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95300954.  
RA NAKAJIMA N., ABE K.;  
RT "Genomic structure of the mouse A-type lamin gene locus encoding somatic and germ cell-specific lamins.";  
RL FEBS LETT. 365:108-114(1995).  
RN [2]  
RP SEQUENCE OF 1-568 FROM N.A.  
RX MEDLINE; 89247440.  
RA RIEDEL W., WERNER D.;  
RT "Nucleotide sequence of the full-length mouse lamin C cDNA and its deduced amino-acid sequence.";  
RL BIOCHIM. BIOPHYS. ACTA 1008:119-122(1989).  
RN [3]  
RP SEQUENCE OF 235-665 FROM N.A.  
RX MEDLINE; 93144345.  
RA NAKAJIMA N., SADO T.;  
RT "Nucleotide sequence of a mouse lamin A cDNA and its deduced amino acid sequence.";  
RL BIOCHIM. BIOPHYS. ACTA 1171:311-314(1993).  
CC -!- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,

CC WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE  
 CC AND MAY ALSO INTERACT WITH CHROMATIN.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- PTM: INCREASED PHOSPHORYLATION OF THE LAMINS OCCURS BEFORE  
 CC ENVELOPE DISINTEGRATION AND PROBABLY PLAYS A ROLE IN REGULATING  
 CC LAMIN ASSOCIATIONS.  
 CC -1- THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY CONTROLLED BY  
 CC THE CELL CYCLE, AS SEEN BY THE DISINTEGRATION AND FORMATION OF  
 CC THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE, RESPECTIVELY.  
 CC -1- ALTERNATIVE PRODUCTS: LAMINS A, C AND C2 ARE THE PRODUCTS OF  
 CC ALTERNATE SPLICING OF THE SAME GENE. LAMINS A AND C ARE PRESENT IN  
 CC EQUAL AMOUNTS IN THE LAMINA OF MAMMALS.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 582  
 CC ONWARD AND IS LONGER (702 AA) DUE TO A FRAMESHIFT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D49733; G1794159; JOINED.  
 CC EMBL: D49732; G1794159; JOINED.  
 CC EMBL: D49734; G1794159; JOINED.  
 CC EMBL: D49735; G1794159; JOINED.  
 CC EMBL: D49732; E212319; JOINED.  
 CC EMBL: D13181; G220474; JOINED.  
 CC MGD; MG1-96794; LMNA.  
 CC PROSITE: PS00226; IF\_1; 1.  
 CC PFAM: PF00038; filament; 1.  
 CC PFAM: PF00932; IF\_tail; 1.  
 CC KW INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL;  
 CC NUCLEAR PROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING.  
 CC FT DOMAIN 1 33 HEAD.  
 CC FT DOMAIN 34 383 ROD.  
 CC FT DOMAIN 384 665 TAIL.  
 CC FT DOMAIN 34 70 COIL 1A.  
 CC FT DOMAIN 71 80 LINKER 1.  
 CC FT DOMAIN 81 218 COIL 1B.  
 CC FT DOMAIN 219 242 LINKER 2.  
 CC FT DOMAIN 243 383 COIL 2.  
 CC FT SITE 325 325 STUTTER (BY SIMILARITY).  
 CC FT LIPID 417 422 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT FARNESYL 662 662 FARNESYL (BY SIMILARITY).  
 CC FT R -> VC (IN REF. 2).  
 CC FT R -> P (IN REF. 2).  
 CC FT RV -> WL (IN REF. 2).  
 CC FT A -> V (IN REF. 1).  
 CC FT A -> V (IN REF. 1).  
 CC SQ SEQUENCE 665 AA; 74209 MW; B4485A65 CRC32;  
 CC -----  
 CC Query Match 78.6%; Score 44; DB 1; Length 665;  
 CC Best Local Similarity 62.5%; Pred. No. 1.16e+01;  
 CC Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 CC -----  
 CC Db 56 NAGLRLRI 63  
 CC :|||:|:  
 CC QY 57 SAGLHLRV 64  
 CC :|||:|:  
 CC -----  
 CC RESULT 9  
 CC ID LAMA-RAT STANDARD; PRT; 665 AA.  
 CC AC P48679;  
 CC DT 01-FEB-1996 (REL. 33, CREATED)  
 CC DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 CC DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 CC DE LAMIN A.  
 CC GN LMN1 OR LMNA.  
 CC OS RATTUS NORVEGICUS (RAT).  
 CC OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 CC OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 93050186.  
 RA OZAKI T., SAKIYAMA S.;  
 RT "Lamin A gene expression is specifically suppressed in  
 RL v-src-transformed cells.";  
 RL FEBS LETT. 312:165-168(1992).  
 RN [2]  
 RP SEQUENCE OF 26-663 FROM N.A.  
 RC JONNALAGADDA V.S. PARNAIK V.K.;  
 RL SUBMITTED (NOV-1993) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS  
 CC LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,  
 CC WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE  
 CC AND MAY ALSO INTERACT WITH CHROMATIN.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- PTM: INCREASED PHOSPHORYLATION OF THE LAMINS OCCURS BEFORE  
 CC ENVELOPE DISINTEGRATION AND PROBABLY PLAYS A ROLE IN REGULATING  
 CC LAMIN ASSOCIATIONS.  
 CC -1- THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY CONTROLLED BY  
 CC THE CELL CYCLE, AS SEEN BY THE DISINTEGRATION AND FORMATION OF  
 CC THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE, RESPECTIVELY.  
 CC -1- ALTERNATIVE PRODUCTS: LAMINS A AND C ARE THE PRODUCTS OF ALTERNATE  
 CC SPLICING OF THE SAME GENE. THEY ARE PRESENT IN EQUAL AMOUNTS IN  
 CC THE LAMINA OF MAMMALS.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN IN THE N- AND  
 CC C-TERMINAL DUE TO FRAMESHIFTS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X66870; G56551; ALT\_FRAME.  
 CC EMBL: X76297; G453180; -.  
 CC PROSITE: PS00226; IF\_1; 1.  
 CC PFAM: PF00038; filament; 1.  
 CC PFAM: PF00932; IF\_tail; 1.  
 CC KW INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL;  
 CC NUCLEAR PROTEIN; LIPID PROTEIN; PRENYLATION; PHOSPHORYLATION;  
 CC ALTERNATIVE SPLICING.  
 CC FT DOMAIN 1 33 HEAD.  
 CC FT DOMAIN 34 383 ROD.  
 CC FT DOMAIN 384 665 TAIL.  
 CC FT DOMAIN 34 70 COIL 1A.  
 CC FT DOMAIN 71 80 LINKER 1.  
 CC FT DOMAIN 81 218 COIL 1B.  
 CC FT DOMAIN 219 242 LINKER 2.  
 CC FT DOMAIN 243 383 COIL 2.  
 CC FT SITE 325 325 STUTTER (BY SIMILARITY).  
 CC FT LIPID 417 422 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT FARNESYL 662 662 FARNESYL (BY SIMILARITY).  
 CC FT K -> R (IN REF. 1).  
 CC FT T -> S (IN REF. 1).  
 CC FT R -> P (IN REF. 1).  
 CC FT CONFLICT 584 584  
 CC FT CONFLICT 606 606 A -> P (IN REF. 1).  
 CC SQ SEQUENCE 665 AA; 74323 MW; 3A1BAC8D CRC32;  
 CC -----  
 CC Query Match 78.6%; Score 44; DB 1; Length 665;  
 CC Best Local Similarity 62.5%; Pred. No. 1.16e+01;  
 CC Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 CC -----  
 CC Db 56 NAGLRLRI 63  
 CC :|||:|:  
 CC QY 57 SAGLHLRV 64  
 CC :|||:|:  
 CC -----  
 CC RESULT 10



```

ID PHT2_PSEPU STANDARD; PRT; 324 AA.
AC Q05182;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PHTHALATE 4,5-DIOXYGENASE OXYGENASE REDUCTASE SUBUNIT (EC 1.14.12.7).
GN PHT2.
OS PSEUDOMONAS PUTIDA.
OG PLASMID PHT.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC PSEUDOMONAS.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-NH102-2.
RA NOMURA Y., NAKAGAWA M., OGAWA N., HARASHIMA S., OSHIMA Y.;
RT "Genes in PHT plasmid encoding the initial degradation pathway of
phthalate in Pseudomonas putida.";
RL J. FERMENT. BIOENG. 74:333-344(1992).
CC -1- CATALYTIC ACTIVITY: PHTHALATE + NADH + O(2) -> CIS-4,5-
DIHYDROXYCYCLOHEXA-1(6),2-DIENE-1,2-DICARBOXYLATE + NAD(+) +
H(2)O.
CC -1- PATHWAY: FIRST STEP OF PHTHALATE DEGRADATION.
CC -1- COFACTOR: FMN (BY SIMILARITY).
CC -1- SUBUNIT: THIS DIOXYGENASE SYSTEM CONSISTS OF TWO PROTEINS:
PHTHALATE OXYGENASE AND PHTHALATE OXYGENASE REDUCTASE.
CC -1- INDUCTION: INDUCED BY PHTHALATE AND REPRESSED BY GLUCOSE.
CC -1- SIMILARITY: IN THE C-TERMINAL, BELONGS TO THE 2FE2S PLANT-TYPE
FERREDOXIN FAMILY.
CC -1- SIMILARITY: BELONGS TO THE PDR/VANB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13229; G295709;
DR PROSITE; PS00197; 2FE2S-FERREDOXIN; 1.
DR PFAM; PF00111; fer2; 1.
DR PFAM; PF00175; oxidored_fad; 1.
DR PLASMID; AROMATIC HYDROCARBONS CATABOLISM; IRON-SULFUR; NAD;
KW OXIDOREDUCTASE; ELECTRON TRANSPORT; MONOOXYGENASE; FMN; FLAVOPROTEIN.
FT NP_BIND 1 105 FMN (BY SIMILARITY).
FT NP_BIND 115 229 NAD (BY SIMILARITY).
FT METAL 275 275 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 280 280 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 283 283 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 311 311 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SQ SEQUENCE 324 AA; 35999 MW; 216AA92F CRC32;

Query Match 76.8%; Score 43; DB 1; Length 324;
Best Local Similarity 75.0%; Pred. No. 1.94e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 43 TAGAHLRV 50
:|:|:|:|
QY 57 SAGLHLRV 64

RESULT 11
ID RP55_RHIME STANDARD; PRT; 513 AA.
AC P33985;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE RNA POLYMERASE SIGMA-54 FACTOR (VERSION 2).
GN RPON OR NTRA.
OS RHIZOBIUM MELILOTI.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; SINORHIZOBIUM.
RN [1]

Query Match 76.8%; Score 43; DB 1; Length 513;
Best Local Similarity 85.7%; Pred. No. 1.94e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 4 SASLHLR 10
:|:|:|:|
QY 57 SAGLHLR 63

RESULT 12
ID RP54_RHIME STANDARD; PRT; 523 AA.
AC P17263;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE RNA POLYMERASE SIGMA-54 FACTOR (VERSION 1).
GN RPON OR NTRA.
OS RHIZOBIUM MELILOTI.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; SINORHIZOBIUM.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 87222158.
RA RONSON C.W., NIXON B.T., ALBRIGHT L.M., AUSUBEL F.M.;
RT "Rhizobium meliloti ntrA (rpon) gene is required for diverse
metabolic functions.";
RL J. BACTERIOL. 169:2424-2431(1987).
CC -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
THEN IS RELEASED.
CC -1- FUNCTION: THE SIGMA FACTOR IS RESPONSIBLE FOR THE EXPRESSION OF
THE NITROGEN FIXATION PROTEINS (NIF OPERON), GLNA AND DCTA FOR

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-104A14;
RX MEDLINE; 89359151.
RA SHATTERS R.G., SOMERVILLE J.E., KAHN M.L.;
RT "Regulation of glutamine synthetase II activity in Rhizobium meliloti
104A14.";
RL J. BACTERIOL. 171:5087-5094(1989).
CC -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
THEN IS RELEASED.
CC -1- FUNCTION: THIS SIGMA FACTOR IS RESPONSIBLE FOR THE EXPRESSION OF
THE NITROGEN FIXATION PROTEINS (NIF OPERON), GLNA AND DCTA FOR
DICARBOXYLATE TRANSPORT. THE OPEN COMPLEX (SIGMA-54 AND CORE RNA
POLYMERASE) SERVES AS THE RECEPTOR FOR RECEIPT OF THE MELTING
SIGNAL FROM THE REMOTELY BOUND ACTIVATOR PROTEINS NIFA, NTRC, OR
DCTD FOR THE EXPRESSION OF THE REGULATED PROTEINS.
CC -1- SIMILARITY: BELONGS TO THE SIGMA-54 FACTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M28846; G152147;
DR PIR; A33966; A33966.
DR PROSITE; PS00717; SIGMA54_1; 1.
DR PROSITE; PS00718; SIGMA54_2; 1.
DR PROSITE; PS00719; SIGMA54_3; 1.
DR PFAM; PF00309; Sigma54_factors; 1.
KW TRANSCRIPTION REGULATION; SIGMA FACTOR; DNA-DIRECTED RNA POLYMERASE;
FT DOMAIN 11 29 GLN-RICH.
FT DOMAIN 21 42 LEUCINE-ZIPPER (POTENTIAL).
FT DOMAIN 190 211 LEUCINE-ZIPPER (POTENTIAL).
FT DNA_BIND 394 413 H-T-H MOTIF (POTENTIAL).
FT SITE 483 491 RPON BOX.
SQ SEQUENCE 513 AA; 56821 MW; F3DE59C4 CRC32;

Query Match 76.8%; Score 43; DB 1; Length 513;
Best Local Similarity 85.7%; Pred. No. 1.94e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 4 SASLHLR 10
:|:|:|:|
QY 57 SAGLHLR 63

RESULT 12
ID RP54_RHIME STANDARD; PRT; 523 AA.
AC P17263;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE RNA POLYMERASE SIGMA-54 FACTOR (VERSION 1).
GN RPON OR NTRA.
OS RHIZOBIUM MELILOTI.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; SINORHIZOBIUM.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 87222158.
RA RONSON C.W., NIXON B.T., ALBRIGHT L.M., AUSUBEL F.M.;
RT "Rhizobium meliloti ntrA (rpon) gene is required for diverse
metabolic functions.";
RL J. BACTERIOL. 169:2424-2431(1987).
CC -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
THEN IS RELEASED.
CC -1- FUNCTION: THE SIGMA FACTOR IS RESPONSIBLE FOR THE EXPRESSION OF
THE NITROGEN FIXATION PROTEINS (NIF OPERON), GLNA AND DCTA FOR

```

CC DICARBOXYLATE TRANSPORT. THE OPEN COMPLEX (SIGMA-54 AND CORE RNA  
CC POLYMERASE) SERVES AS THE RECEPTOR FOR RECEIPT OF THE MELTING  
CC SIGNAL FROM THE REMOTELY BOUND ACTIVATOR PROTEINS NIFA, NTRC, OR  
CC DCTD FOR THE EXPRESSION OF THE REGULATED PROTEINS.  
CC -1- SIMILARITY: BELONGS TO THE SIGMA-54 FACTOR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M16513; G152390; -  
DR PROSITE: PS00717; SIGMA54\_1; 1.  
DR PROSITE: PS00718; SIGMA54\_2; 1.  
DR PROSITE: PS50044; SIGMA54\_3; 1.  
DR PFAM: PF00309; Sigma54\_factors; 1.  
KW TRANSCRIPTION REGULATION; SIGMA FACTOR; DNA-DIRECTED RNA POLYMERASE;  
KW DNA-BINDING; NITROGEN FIXATION.  
FT DOMAIN 11 29 GLN-RICH  
FT DOMAIN 21 42 LEUCINE-ZIPPER (POTENTIAL).  
FT DOMAIN 189 210 LEUCINE-ZIPPER (POTENTIAL).  
FT DNA\_BIND 393 412 H-T-H MOTIF (POTENTIAL).  
FT SITE 482 490 RPN BOX.  
SQ SEQUENCE 523 AA; 57815 MW; 2EB5FCD7 CRC32;  
  
Query Match 76.8%; Score 43; DB 1; Length 523;  
Best Local Similarity 85.7%; Pred. No. 1.94e+01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Db 4 SASLHLR 10  
QY 57 SAGLHLR 63  
II:IIII  
  
RESULT 13  
ID RP54\_RHISN STANDARD; PRT; 525 AA.  
AC P22881;  
DT 01-AUG-1991 (REL. 19, CREATED)  
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
DE RNA POLYMERASE SIGMA-54 FACTOR.  
GN RPN OR NTR.  
OS RHIZOBIUM SP. (STRAIN NGR234).  
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;  
OC RHIZOBIACEAE; RHIZOBIUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 91008923.  
RA VAN SLOOTEN J.C., CERVANTES E., BROUGHTON W.J., WONG C.H., STANLEY J.;  
RT "Sequence and analysis of the rpn sigma factor gene of rhizobium sp.  
RT strain NGR234, a primary coregulator of symbiosis."  
RL J. BACTERIOL. 172:5563-5574(1990).  
CC -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES  
CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND  
CC THEN IS RELEASED.  
CC -1- FUNCTION: THIS SIGMA FACTOR IS RESPONSIBLE FOR THE EXPRESSION OF  
CC THE NITROGEN FIXATION PROTEINS (NIF OPERON), GLNA, AND DCTA FOR  
CC DICARBOXYLATE TRANSPORT. THE OPEN COMPLEX (SIGMA-54 AND CORE RNA  
CC POLYMERASE) SERVES AS THE RECEPTOR FOR RECEIPT OF THE MELTING  
CC SIGNAL FROM THE REMOTELY BOUND ACTIVATOR PROTEINS NIFA, NTRC, OR  
CC DCTD FOR THE EXPRESSION OF THE REGULATED PROTEINS.  
CC -1- SIMILARITY: BELONGS TO THE SIGMA-54 FACTOR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL: M58481; G152432; -  
CC PIR: A36130; A36130.  
CC PROSITE: PS00717; SIGMA54\_1; 1.  
CC PROSITE: PS00718; SIGMA54\_2; 1.  
CC PROSITE: PS50044; SIGMA54\_3; 1.  
CC PFAM: PF00309; Sigma54\_factors; 1.  
KW TRANSCRIPTION REGULATION; SIGMA FACTOR; DNA-DIRECTED RNA POLYMERASE;  
KW DNA-BINDING. 11 29 GLN-RICH  
FT DOMAIN 21 42 LEUCINE-ZIPPER (POTENTIAL).  
FT DOMAIN 191 212 LEUCINE-ZIPPER (POTENTIAL).  
FT DNA\_BIND 393 412 H-T-H MOTIF (POTENTIAL).  
FT SITE 484 492 RPN BOX.  
SQ SEQUENCE 525 AA; 57800 MW; 6691BDA1 CRC32;  
  
Query Match 76.8%; Score 43; DB 1; Length 525;  
Best Local Similarity 85.7%; Pred. No. 1.94e+01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Db 4 SASLHLR 10  
QY 57 SAGLHLR 63  
II:IIII  
  
RESULT 14  
ID DSG2\_HUMAN STANDARD; PRT; 1117 AA.  
AC Q14126;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE DESMOGLEIN 2 PRECURSOR (HDGC).  
GN DSG2  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94192736.  
RA SCHAEFER S., KOCH P.J., FRANK W.W.;  
RT "Identification of the ubiquitous human desmoglein, Dsg2, and the  
RT expression catalogue of the desmoglein subfamily of desmosomal  
RT cadherins."  
RL EXP. CELL RES. 211:391-399(1994).  
RN [2]  
RP SEQUENCE OF 777-1117 FROM N.A.  
RX MEDLINE; 92037656.  
RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R., FRANK W.W.;  
RT "Complete amino acid sequence of the epidermal desmoglein precursor  
RT polypeptide and identification of a second type of desmoglein gene."  
RL EUR. J. CELL BIOL. 55:200-208(1991).  
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.  
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE  
CC FILAMENTS MEDIATING CELL-CELL ADHESION.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.  
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
CC (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE  
CC DESMOsome SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z26317; G416178; -  
DR MIM: 125671; -  
DR PROSITE: PS00232; CADHERIN; 3.

DR PFAM; PF00028; cadherin; 4.  
DR HSP; P15116; INCI.  
KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;  
KW CYTOSKELETON; CALCIUM-BINDING.  
FT SIGNAL 1 23 POTENTIAL.  
FT PROPEP 24 48 POTENTIAL.  
FT CHAIN 49 1117 DESMOGLEIN 2.  
FT DOMAIN 49 608 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 609 633 POTENTIAL.  
FT DOMAIN 634 1117 CYTOPLASMIC (POTENTIAL).  
FT REPEAT 49 159 CADHERIN 1.  
FT REPEAT 160 272 CADHERIN 2.  
FT REPEAT 273 387 CADHERIN 3.  
FT REPEAT 388 502 CADHERIN 4.  
FT REPEAT 880 911 DESMOGLEIN REPEAT 1.  
FT REPEAT 912 941 DESMOGLEIN REPEAT 2.  
FT REPEAT 942 967 DESMOGLEIN REPEAT 3.  
FT REPEAT 968 991 DESMOGLEIN REPEAT 4.  
FT REPEAT 992 1020 DESMOGLEIN REPEAT 5.  
FT REPEAT 1021 1050 DESMOGLEIN REPEAT 6.  
FT CARBOHYD 111 111 POTENTIAL.  
FT CARBOHYD 181 181 POTENTIAL.  
FT CARBOHYD 308 308 POTENTIAL.  
FT CARBOHYD 461 461 POTENTIAL.  
FT CARBOHYD 513 513 POTENTIAL.  
SQ SEQUENCE 1117 AA; 122385 MW; 84D3B898 CRC32;

Query Match 76.8%; Score 43; DB 1; Length 1117;  
Best Local Similarity 62.5%; Pred. No. 1.94e+01;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 20 GSGHLQV 27  
:|||||  
QY 57 SAGLHLRV 64

RESULT 15  
ID VIT6\_OSCBR STANDARD; PRT; 1660 AA.  
AC Q94637;  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE VITELLOGENIN 6 PRECURSOR.  
GN VIT-6.  
OS OSCEIUS BREVESOPHAGA.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;  
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; RHABDITINAE; OSCEIUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CEWI;  
RX MEDLINE; 96212989.  
RA WINTER C.E., PENHA C., BLUMENTHAL T.;  
RT "Comparison of a vitellogenin gene between two distantly related  
rhabditid nematode species";  
RL MOL. BIOL. EVOL. 13:674-684(1996).  
CC -!- FUNCTION: PRECURSOR OF THE EGG-YOLK PROTEINS THAT ARE SOURCES OF  
NUTRIENTS DURING EMBRYONIC DEVELOPMENT (POTENTIAL).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U35449; G1515337; -  
DR HSP; P11064; IPR.  
KW STORAGE PROTEIN; MULTIGENE FAMILY; SIGNAL.  
FT SIGNAL 1 15 POTENTIAL.  
FT CHAIN 16 1660 VITELLOGENIN 6.  
FT CARBOHYD 237 237 POTENTIAL.  
FT CARBOHYD 383 383 POTENTIAL.

FT CARBOHYD 695 695 POTENTIAL.  
FT CARBOHYD 1307 1307 POTENTIAL.  
FT CARBOHYD 1596 1596 POTENTIAL.  
FT CARBOHYD 1629 1629 POTENTIAL.  
SQ SEQUENCE 1660 AA; 192108 MW; 9AA62413 CRC32;

Query Match 76.8%; Score 43; DB 1; Length 1660;  
Best Local Similarity 71.4%; Pred. No. 1.94e+01;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 887 TGLRLRV 893  
:|||||  
QY 58 AGLHLRV 64

Search completed: Wed Sep 1 16:26:45 1999  
Job time : 8 secs.

\*\*\*\*\*  
W P S R L H (TW)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Sep 1 16:27:02 1999; MasPar time 5.10 Seconds  
Tabular output not generated. 85.590 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (57-64) from PCTUS9913024.pep (10 of 12)  
Perfect Score: 56  
Sequence: 1 SAGLHLRV 8

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl9

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle  
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp\_unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 21.702; Variance 24.925; scale 0.871

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description	Pred. No.
1	50	89.3	4293	11	O08852	POLYCYSTIC KIDNEY DISE	1.66e+00
2	48	85.7	415	10	O08705	F8K4.20 PROTEIN.	4.68e+00
3	48	85.7	479	5	O17697	SIMILAR TO GLUTAMATE R	4.68e+00
4	48	85.7	2605	2	Q50858	SAFRAMYCIN MX1 SYNTHET	4.68e+00
5	47	83.9	191	2	P97168	INTEGRASE (FRAGMENT).	7.77e+00
6	47	83.9	284	2	O05288	INTEGRASE (FRAGMENT).	7.77e+00
7	47	83.9	303	2	Q49186	INTEGRASE.	7.77e+00
8	47	83.9	336	2	Q51431	INTEGRASE (FRAGMENT).	7.77e+00
9	47	83.9	337	2	Q56438	INTEGRASE.	7.77e+00
10	47	83.9	337	2	Q56440	HYPOTHETICAL 38.4 KD P	7.77e+00
11	47	83.9	337	2	Q48337	INTEGRASE.	7.77e+00
12	47	83.9	337	2	O50324	INTEGRASE.	7.77e+00
13	45	80.4	238	14	P90337	REPLICATION-RELATED PR	2.10e+01
14	45	80.4	260	2	O68996	HYPOTHETICAL 27.8 KD P	2.10e+01
15	45	80.4	499	2	O07561	HYPOTHETICAL 54.4 KD P	2.10e+01
16	45	80.4	755	14	P89107	SCVP86.	2.10e+01
17	45	80.4	2133	14	O98203	MC035R.	2.10e+01
18	44	78.6	438	8	O80030	RIBULOSE BISPSPHOSPHATE	3.41e+01
19	44	78.6	440	8	Q32141	RIBULOSE BISPSPHOSPHATE	3.41e+01
20	44	78.6	440	8	O78589	RIBULOSE BISPSPHOSPHATE	3.41e+01

21	44	78.6	447	8	O78577	RIBULOSE BISPSPHOSPHATE	3.41e+01
22	44	78.6	447	8	Q37324	RIBULOSE BISPSPHOSPHATE	3.41e+01
23	44	78.6	447	8	O96083	RIBULOSE BISPSPHOSPHATE	3.41e+01
24	44	78.6	466	8	P93933	RIBULOSE BISPSPHOSPHATE	3.41e+01
25	44	78.6	467	8	Q37330	RIBULOSE BISPSPHOSPHATE	3.41e+01
26	43	76.8	341	1	O29051	CONSERVED HYPOTHETICAL	5.49e+01
27	43	76.8	490	5	O19324	CODED FOR BY C. ELEGAN	5.49e+01
28	43	76.8	496	5	Q27693	CYTCHROME P450 (EC 1.	5.49e+01
29	43	76.8	514	2	O59751	RNA POLYMERASE SIGMA-5	5.49e+01
30	43	75.8	884	10	O04085	UNKNOWN PROTEIN.	5.49e+01
31	42	75.0	170	2	O52584	HYPOTHETICAL 17.7 KD P	8.78e+01
32	42	75.0	216	8	O48269	CYTCHROME C OXIDASE S	8.78e+01
33	42	75.0	321	4	O99608	NECDIN RELATED PROTEIN	8.78e+01
34	42	75.0	409	14	O64961	NUCLEOCAPSID PROTEIN.	8.78e+01
35	42	75.0	442	13	O73803	FUGU HEDGEHOG.	8.78e+01
36	42	75.0	453	4	O14586	BAC CLONE GS303P24 FRO	8.78e+01
37	42	75.0	494	10	Q39520	LEGUMIN (FRAGMENT).	8.78e+01
38	42	75.0	502	10	O40689	LEGUMIN.	8.78e+01
39	42	75.0	561	3	O60051	COHESIN.	8.78e+01
40	42	75.0	689	13	O13265	TISSUE TRANSGLUTAMINAS	8.78e+01
41	42	75.0	713	3	O42947	HYPOTHETICAL 80.1 KD P	8.78e+01
42	42	75.0	849	5	O77068	ALPHA ACTININ (FRAGMEN	8.78e+01
43	42	75.0	1072	6	O77690	OLFACTORY ENTEROTOXIN	8.78e+01
44	42	75.0	1262	11	Q60988	TALI INTERRUPTING LOCUS	8.78e+01
45	42	75.0	1706	11	Q63755	RETINOBLASTOMA PROTEIN	8.78e+01

ALIGNMENTS

RESULT 1  
ID O08852 PRELIMINARY; PRT: 4293 AA.  
AC O08852;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE POLYCYSTIC KIDNEY DISEASE 1 PROTEIN.  
GN PKD1.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCUROGNATHI; MURIDAE; MURINAE; MUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97262094.  
RA LOHNING C., NOWICKA U., FRISCHAUF A.M.;  
RT "The mouse homolog of PKD1: sequence analysis and alternative  
RT splicing."  
RL MAMM. GENOME 8:307-311(1997).  
DR ENBL; U70209; G2138183; -.  
DR PFAM; PF00560; LRR; 1.  
DR PFAM; PF00801; PKD; 16.  
SQ SEQUENCE 4293 AA; 466545 MW; 10E37A8A CRC32;

Query Match 89.3%; Score 50; DB 11; Length 4293;  
Best Local Similarity 85.7%; Pred. No. 1.66e+00;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 2910 AGLHLRI 2916  
Qy 58 AGLHLRV 64

RESULT 2  
ID O08705 PRELIMINARY; PRT: 415 AA.  
AC O08705;  
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE F8K4.20 PROTEIN.  
GN F8K4.20.

OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; TRACHEOPHYTA;  
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;  
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA VYOTSKAIA V.S., SCHWARTZ J.R., TORIUMI M., KWAN A., YU G., OJI O.,
LIU S., LI J., ARAUJO R., AU M., BRENDEN V., BUEHLER E., CONWAY A.B.,
CONWAY A.R., DEWAR K., FENG J., KIM C., KURTZ D., LI Y., PALM C.J.,
SHINN P., SUN H., DAVIS R.W., ECKER J.R., FEDERSPIEL N.A.,
RA THEOLOGIS A.;
RT "Arabidopsis thaliana chromosome 1 BAC F8K4 sequence.";
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA THEOLOGIS A.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA THEOLOGIS A.;
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA THEOLOGIS;
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AC004392; G3367531;
SQ SEQUENCE 415 AA; 46438 MW; B5E0FFEL CRC32;
Query Match 85.7%; Score 48; DB 10; Length 415;
Best Local Similarity 100.0%; Pred. No. 4.68e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 14 GLHLRV 19
QY 59 GLHLRV 64
RESULT 3
ID Q17697 PRELIMINARY; PRT; 479 AA.
AC Q17697;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO GLUTAMATE RECEPTORS.
GN COB48.9.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
BOFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA LEIMBACH D.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
```

```
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U39849; G1055049;
DR PFAM: PF00060; 11g_chan; 2.
SQ SEQUENCE 479 AA; 54976 MW; 604C46C9 CRC32;
Query Match 85.7%; Score 48; DB 5; Length 479;
Best Local Similarity 100.0%; Pred. No. 4.68e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 46 GLHLRV 51
QY 59 GLHLRV 64
RESULT 4
ID Q50858 PRELIMINARY; PRT; 2605 AA.
AC Q50858;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE SAFRAMYCIN MX1 SYNTHETASE A.
GN SAFA.
OS MYXOCOCCUS XANTHUS.
OC BACTERIA; PROTEOBACTERIA; DELTA SUBDIVISION; MYXOBACTERIA;
OC MYXOCOCCALES; CYSTOBACTERINEAE; MYXOCOCCACEAE; MYXOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DM504-15;
RX MEDLINE; 96032006.
RA POSPIECH A., CLUZEL B., BIETENHADER J., SCHUPP T.;
RT "A new Myxococcus xanthus gene cluster for the biosynthesis of the
antibiotic saframycin Mx1 encoding a peptide synthetase.";
RL MICROBIOLOGY 141:1793-1803(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-DM504-15;
RX MEDLINE; 97090395.
RA POSPIECH A., BIETENHADER J., SCHUPP T.;
RT "two multifunctional peptide synthetases and an O-methyltransferase
are involved in the biosynthesis of the DNA-binding antibiotic and
antitumour agent saframycin Mx1 from Myxococcus xanthus.";
RL MICROBIOLOGY 142:741-746(1996).
DR EMBL: U24657; G1171129;
DR PROSITE; PS00455; AMP_BINDING; 2.
DR PFAM: PF00501; AMP-binding; 2.
DR PFAM: PF00550; pp-binding; 2.
DR PFAM: PF00668; DUF4; 2.
SQ SEQUENCE 2605 AA; 285745 MW; 09C34B84 CRC32;
Query Match 85.7%; Score 48; DB 2; Length 2605;
Best Local Similarity 100.0%; Pred. No. 4.68e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1530 GLHLRV 1535
QY 59 GLHLRV 64
RESULT 5
ID P97168 PRELIMINARY; PRT; 191 AA.
AC P97168;
DT 01-MAY-1997 (TREMREL. 03, CREATED)
DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE INTEGRASE (FRAGMENT).
GN INI.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
```

RP SEQUENCE FROM N.A.  
RC STRAIN-RA33.2;  
RA ADRIAN P.V., THOMSON C.J., KLUGMAN K.P., AMYES S.G.B.;  
RL SUBMITTED (AUG-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; 250802; G950043;  
FT NON\_TER 191  
SQ SEQUENCE 191 AA; 21891 MW; C007CICE CRC32;

Query Match 83.9%; Score 47; DB 2; Length 191;  
Best Local Similarity 75.0%; Pred. No. 7.77e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLQVR 155  
QY 57 SAGLHLRV 64

RESULT 6  
ID O05288 PRELIMINARY; PRT; 284 AA.  
AC O05288;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE INTEGRASE (FRAGMENT).  
OS KLEBSIELLA OXYTOCA.  
OG PLASMID PACM1.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
OC KLEBSIELLA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97312761.  
RA PRESTON K.E.; KACICA M.A., LIMBERGER R.J., ARCHINAL W.A.,  
RA VENEZIA R.A.;  
RT "The resistance and integrase genes of PACM1, a conjugative  
multiple-resistance plasmid, from Klebsiella oxytoca";  
RL PLASMID 37:105-118(1997).  
DR EMBL; U09445; G1906652;  
DR PFAM; PF00589; Phage\_integrase; 1.  
KW PLASMID.  
FT NON\_TER 284  
SQ SEQUENCE 284 AA; 32632 MW; 6DSALB7F CRC32;

Query Match 83.9%; Score 47; DB 2; Length 284;  
Best Local Similarity 75.0%; Pred. No. 7.77e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLQVR 155  
QY 57 SAGLHLRV 64

RESULT 7  
ID Q49186 PRELIMINARY; PRT; 303 AA.  
AC Q49186;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE INTEGRASE.  
OS MYCOBACTERIUM FORTUITUM.  
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIACEAE;  
OC ACTINOMYCETALES; CORYNEBACTERIACEAE; MYCOBACTERIUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSOM-TN10;  
RX MEDLINE; 90294910.  
RA MARTIN C., TIMM J., RAUZIER J., GOMEZ-LUS R., DAVIES J., GICQUEL B.;  
RT "Transposition of an antibiotic resistance element in mycobacteria";  
RL NATURE 345:739-743(1990).  
DR EMBL; X53635; G4286;  
DR PFAM; PF00589; Phage\_integrase; 1.  
SQ SEQUENCE 303 AA; 34884 MW; IC45D6BB CRC32;

Query Match 83.9%; Score 47; DB 2; Length 303;

Best Local Similarity 75.0%; Pred. No. 7.77e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Db 148 SEGQLQVR 155  
QY 57 SAGLHLRV 64

RESULT 8  
ID Q51431 PRELIMINARY; PRT; 336 AA.  
AC Q51431;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE INTEGRASE (FRAGMENT).  
GN TNPI.  
OS PSEUDOMONAS AERUGINOSA.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;  
OC PSEUDOMONAS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-101;  
RA LARAKI N., ROSSOLINI G., GALLEN M., FRERE J.M.;  
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; X98393; E248569;  
DR PFAM; PF00589; Phage\_integrase; 1.  
FT NON\_TER 336  
SQ SEQUENCE 336 AA; 38239 MW; 420D38E3 CRC32;

Query Match 83.9%; Score 47; DB 2; Length 336;  
Best Local Similarity 75.0%; Pred. No. 7.77e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLQVR 155  
QY 57 SAGLHLRV 64

RESULT 9  
ID Q56438 PRELIMINARY; PRT; 337 AA.  
AC Q56438;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE INTEGRASE.  
GN TNPI.  
OS SHIGELLA SONNEI.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
OC SHIGELLA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSOM-TN2426;  
RX MEDLINE; 86041899.  
RA FLING M.E., KOFF J., RICHARDS C.;  
RT "Nucleotide sequence of the transposon Tn7 gene encoding an  
aminoglycoside-modifying enzyme, 3'(9)-O-nucleotidyltransferase";  
RL NUCLEIC ACIDS RES. 13:7095-7106(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSOM-TN2426;  
RA WIEDEMANN B.;  
RL ANTIMICROB. AGENTS CHEMOTHER. 0:85-92(1986).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSOM-TN2426;  
RX MEDLINE; 87066734.  
RA CAMERON F.H., GROOT OBBINK D.J., ACKERMAN V.P., HALL R.M.;  
RT "Nucleotide sequence of the AAD(2'') aminoglycoside  
adenylyltransferase determinant aadB. Evolutionary relationship of  
this region with those surrounding aadA in R538-1 and dhfrII in  
R388";  
RL NUCLEIC ACIDS RES. 14:8625-8635(1986).  
RN [4]

```

RP SEQUENCE FROM N.A.
RC TRANSPOSON-TN2426;
RX MEDLINE: 88086917.
RA TENOVER F.C., FILPULA D., PHILLIPS K.L., FLORDE J.J.;
RT "Cloning and sequencing of a gene encoding an aminoglycoside
RL 6'-N-acetyltransferase from an R factor of Citrobacter diversus.";
RN J. BACTERIOL. 170:471-473(1988).
RN [5]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-TN2426;
RX MEDLINE: 90299796.
RA MERCIER J., LACHAPPELLE J., COUTURE F., LAFOND M., VEZINA G.,
RT BOISSINOT M., LEVESQUE R.C.;
RT "Structural and functional characterization of tnpI, a recombinase
RL locus in Tn21 and related beta-lactamase transposons.";
RN J. BACTERIOL. 172:3745-3757(1990).
RN [6]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-TN2426;
RX MEDLINE: 93329367.
RA ZUHLSDORF M.T., WIEDEMANN B.;
RT "Functional and physiological characterization of the Tn21 cassette
RL for resistance genes in Tn2426.";
RN J. GEN. MICROBIOL. 139:995-1002(1993).
DR EMBL; M86913; G155024; -.
DR PFAM; PF00589; Phage_integrase; 1.
SQ SEQUENCE 337 AA; 38415 MW; 093D245D CRC32;

Query Match      83.9%; Score 47; DB 2; Length 337;
Best Local Similarity 75.0%; Pred. No. 7.77e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155
QY 57 SAGLHLRV 64

RESULT 10
ID Q56440 PRELIMINARY; PRT; 337 AA.
AC Q56440;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 38.4 KD PROTEIN.
OS UNKNOWN.
OC BACTERIA; ENVIRONMENTAL SAMPLES.
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-TN21;
RX MEDLINE: 90299796.
RA MERCIER J., LACHAPPELLE J., COUTURE F., LAFOND M., VEZINA G.,
RT BOISSINOT M., LEVESQUE R.C.;
RT "Structural and functional characterization of tnpI, a recombinase
RL locus in Tn21 and related beta-lactamase transposons.";
RN J. BACTERIOL. 172:3745-3757(1990).
DR EMBL; M33633; G1197009; -.
DR PFAM; PF00589; Phage_integrase; 1.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 337 AA; 38376 MW; D8B33A67 CRC32;

Query Match      83.9%; Score 47; DB 2; Length 337;
Best Local Similarity 75.0%; Pred. No. 7.77e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155
QY 57 SAGLHLRV 64

RESULT 11
ID Q48337 PRELIMINARY; PRT; 337 AA.
AC Q48337; P71194; Q48386;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

```

```

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE INTEGRASE.
GN INT.
OS ESCHERICHIA COLI, AND KLEBSIELLA AEROGENES.
OC PLASMID PDG0100, AND PLASMID INCP-BETA R751.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87066734.
RA CAMERON F.H., GROOT OBBINK D.J., ACKERMAN V.P., HALL R.M.;
RT "Nucleotide sequence of the AAD(2'') aminoglycoside
RT adenylyltransferase determinant aadB. Evolutionary relationship of
RT this region with those surrounding aadA in R538-1 and dhfrII in
RT R388.";
RL NUCLEIC ACIDS RES. 14:8625-8635(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88015610.
RA HALL R.M., VOCKLER C.;
RT "The region of the IncN plasmid R46 coding for resistance to
RT beta-lactam antibiotics, streptomycin/spectinomycin and sulphonamides
RT is closely related to antibiotic resistance segments found in IncW
RT plasmids and in Tn21-like transposons.";
RL NUCLEIC ACIDS RES. 15:7491-7501(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90272804.
RA HALL R.M., STOKES H.W.;
RT "The structure of a partial duplication in the integron of plasmid
RT PDG0100.";
RL PLASMID 23:76-79(1990).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92206862.
RA PARSONS Y., HALL R.M., STOKES H.W.;
RT "A new trimethoprim resistance gene, dhfrX, in the In7 integron of
RT plasmid PDG0100.";
RL ANTIMICROB. AGENTS CHEMOTHER. 35:2436-2439(1991).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93391548.
RA STOKES H.W., TOMARAS C., PARSONS Y., HALL R.M.;
RT "The partial 3'-conserved segment duplications in the integrons In6
RT from pSa and In7 from PDG0100 have a common origin.";
RL PLASMID 30:39-50(1993).
RN [6]
RP SEQUENCE FROM N.A.
RC SPECIES-AEROBACTER AEROGENES;
RX MEDLINE: 85215496.
RA SMITH C.A., THOMAS C.M.;
RT "Comparison of the nucleotide sequences of the vegetative replication
RT origins of broad host range Incp plasmids R751 and RK2 reveals
RT conserved features of probable functional importance.";
RN [7]
RP SEQUENCE FROM N.A.
RA SUNDBLOM L.;
RL SUBMITTED (MAR-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [8]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94252994.
RA RADSTROM P., SKOLD O., SWEDBERG G., FLENSBURG J., ROY P.H.,
RT SUNDBLOM L.;
RT "Transposon Tn5090 of plasmid R751, which carries an integron, is
RT related to Tn7, Mu, and the retroelements.";
RN J. BACTERIOL. 176:3257-3268(1994).
DR EMBL; L06418; G149117; -.
DR EMBL; U67194; G1572565; -.
DR EMBL; X72585; G288634; -.
DR PFAM; PF00589; Phage_integrase; 1.
KW PLASMID.

```

SQ SEQUENCE 337 AA; 38358 MW; D25371F2 CRC32;

Query Match 83.9%; Score 47; DB 2; Length 337;  
Best Local Similarity 75.0%; Pred. No. 7.77e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155  
QY 57 SAGLHLRV 64

RESULT 12  
ID O50324 PRELIMINARY; PRT: 337 AA.  
AC O50324;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE INTEGRASE.  
GN INT.  
OS PSEUDOMONAS AERUGINOSA.  
OG PLASMID PMS350.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;  
OC PSEUDOMONAS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PLASMID-PMS350;  
RX MEDLINE: 97176093.  
RA IYOBE S., MINAMI S., YAMADA H.;  
RT Insertion of a carboxypeptidase gene cassette into an integron of a  
RT Pseudomonas aeruginosa plasmid[published erratum appears in J  
RT Antimicrob Chemother 1997 Jun;39(6):845].  
RL J. ANTIMICROB. CHEMOTHER. 38:1114-1115(1996).  
DR EMBL: D78374; D1024672; -.  
SQ SEQUENCE 337 AA; 38298 MW; 753AD743 CRC32;

Query Match 83.9%; Score 47; DB 2; Length 337;  
Best Local Similarity 75.0%; Pred. No. 7.77e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155  
QY 57 SAGLHLRV 64

RESULT 13  
ID P90337 PRELIMINARY; PRT: 238 AA.  
AC P90337;  
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)  
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE REPLICATION-RELATED PROTEIN.  
GN SCVP26.  
OS SAGUARO SARRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; TOMBUSVIRIDAE;  
OC VIRUSES; CACTUS VIRUS.  
OC CARMOVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA XIONG Z., WENG Z.;  
RL J. GEN. VIROL. 0:0-0(0).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA XIONG Z., WENG Z.;  
RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U72332; G1685120; -.  
SQ SEQUENCE 238 AA; 26481 MW; 446D4179 CRC32;

Query Match 80.4%; Score 45; DB 14; Length 238;  
Best Local Similarity 71.4%; Pred. No. 2.10e+01;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 23 AGLQRI 29  
QY 58 AGLHLRV 64

RESULT 14  
ID O68996 PRELIMINARY; PRT: 260 AA.  
AC O68996;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 27.8 KD PROTEIN.  
OS MYCOBACTERIUM AVIUM.  
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;  
OC ACTINOMYCETALES; CORYNEBACTERIACEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-2151;  
RA ECKSTEIN T.M., SILBAO F.S., INAMINE I.M., BELISLE J.T.;  
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AF060182; G3170647; -.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 260 AA; 27809 MW; 882DB702 CRC32;

Query Match 80.4%; Score 45; DB 2; Length 260;  
Best Local Similarity 75.0%; Pred. No. 2.10e+01;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 213 SSGLYLRV 220  
QY 57 SAGLHLRV 64

RESULT 15  
ID O07561 PRELIMINARY; PRT: 499 AA.  
AC O07561;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 54.4 KD PROTEIN.  
GN YHJG.  
OS BACILLUS SUBTILIS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;  
OC BACILLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-168;  
RA NOBACK M.A., TERPSTRA P., HOLSAPPEL S., VENEMA G., BRON S.;  
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 98044033.  
RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.;  
RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,  
RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,  
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,  
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,  
RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.,  
RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,  
RA FRITZ C., FUJITA M., FUJITA Y., FUNA S., GALIZZI A., GALLERON N.,  
RA GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,  
RA GUISEPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,  
RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,  
RA JORIS B., KARAMATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.,  
RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUNANO M.,  
RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,  
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,  
RA MEDINA D., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,  
RA NOONE N., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.,  
RA PARRO V., POHL T.M., PORTETELLE D., PORRILLIK S., PRESCOTT A.M.,  
RA PRESCAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,  
RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADATE Y.,  
RA SATO T., SCANLAN E., SCHLEICH S., SCHROTER R., SCOFFONE F.,  
RA SERIGUCHI J., SEKOWSKA A., SERO S.J., SERROR P., SHIN B.S., SOLDI B.,  
RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,



RA TAKEUCHI M., TAKAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,  
RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,  
RA VIARI A., WAMBUUT R., WEDLER E., WEDLER H., WEITZENEGGER T.,  
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,  
RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,  
RT "The complete genome sequence of the gram-positive bacterium Bacillus  
subtilis.";  
RL NATURE 390:249-256(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;  
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: Y14081; E324980; -;  
DR EMBL: Z99109; E1183052; -;  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 499 AA; 54356 MW; 86F6D7E9 CRC32;  
  
Query Match 80.4%; Score 45; DB 2; Length 499;  
Best Local Similarity 100.0%; Pred. No. 2.10e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 352 AGLHLR 357  
QY 58 AGLHLR 63  
|||||

Search completed: Wed Sep 1 16:27:28 1999  
Job time : 26 secs.

\*\*\*\*\*

W A R E H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:30:31 1999; MasPar time 4.42 Seconds  
67.410 Million cell updates/sec

Tabular output not generated.

Title: >PCT-US99-13024-2  
Description: (66-79) from PCTUS9913024.pep (11 of 12)  
Perfect Score: 86  
Sequence: 1 DPSTGALVDSKSYA 14

Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 17.858; Variance 46.416; scale 0.385

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	52	60.5	50	9	Excitatory amino acid	3.71e+01
2	52	60.5	146	35	Bacterial periplasmic	3.71e+01
3	52	60.5	849	11	Human EAA3d excitator	3.71e+01
4	52	60.5	865	11	Human EAA3c excitator	3.71e+01
5	52	60.5	901	2	Glutamate receptor 6	3.71e+01
6	52	60.5	905	16	Human EAA3 receptor (	3.71e+01
7	52	60.5	905	11	Human EAA3b excitator	3.71e+01
8	52	60.5	905	11	Human EAA3a excitator	3.71e+01
9	52	60.5	908	16	Human EAA4 receptor (	3.71e+01
10	52	60.5	908	9	Excitatory amino acid	3.71e+01
11	52	60.5	919	10	Excitatory amino acid	3.71e+01
12	52	60.5	920	2	Glutamate receptor 5-	3.71e+01
13	51	59.3	50	9	Excitatory amino acid	4.90e+01
14	50	58.1	259	3	Sequence of Oncogene	6.45e+01
15	50	58.1	371	29	N. meningitidis alpha	6.45e+01
16	49	57.0	577	2	Pseudorabies virus gp	8.49e+01

17	49	57.0	577	12	R63144	Glycoprotein I(gpi) o	8.49e+01
18	49	57.0	984	5	R27529	Plasmodium falciparum	8.49e+01
19	49	57.0	984	13	R68037	Plasmodium falciparum	8.49e+01
20	49	57.0	984	1	P80550	Malaria-specific 140k	8.49e+01
21	49	57.0	984	4	R22247	SERP antigen.	8.49e+01
22	49	57.0	984	11	R60164	SERP protein encoded	8.49e+01
23	49	57.0	989	1	R05526	The SERA protein of P	8.49e+01
24	49	57.0	1321	17	R96994	Mouse IRS-2.	8.49e+01
25	48	55.8	28	21	W12039	Curvularia verruculos	1.11e+02
26	48	55.8	469	17	R97242	SCK phosphotyrosine b	1.11e+02
27	48	55.8	490	27	W23247	Human ShcB protein.	1.11e+02
28	48	55.8	600	21	W12042	Curvularia verruculos	1.11e+02
29	48	55.8	899	17	R90987	Nitrate reductase.	1.11e+02
30	47	54.7	144	22	W20516	H. pylori cytoplasmic	1.46e+02
31	47	54.7	303	20	W08956	Chimeric restriction	1.46e+02
32	47	54.7	377	22	W20951	H. pylori secreted or	1.46e+02
33	47	54.7	570	17	R96300	Foki insertion mutant	1.46e+02
34	47	54.7	571	22	W01676	Influenza A/Johannesb	1.46e+02
35	47	54.7	571	22	W01673	Influenza A/Shandong/	1.46e+02
36	47	54.7	571	11	R58759	Foki amino acid seque	1.46e+02
37	47	54.7	579	14	R72810	Foki.	1.46e+02
38	47	54.7	583	1	R06978	Modified enzyme from	1.46e+02
39	47	54.7	585	17	R88595	Foki insertion mutant	1.46e+02
40	47	54.7	589	17	R88596	Foki insertion mutant	1.46e+02
41	47	54.7	596	17	R88597	Foki insertion mutant	1.46e+02
42	47	54.7	600	17	R88598	Foki insertion mutant	1.46e+02
43	47	54.7	2233	33	W48712	HPiV-3 FRhl cp45 vacc	1.46e+02
44	47	54.7	2233	33	W48711	HPiV-3 JS isolate wil	1.46e+02
45	47	54.7	2233	33	W48713	HPiV-3 Vero cp45 vacc	1.46e+02

#### ALIGNMENTS

RESULT 1  
ID R45368 standard; Protein; 50 AA.  
AC R45368.  
DT 07-JUL-1994 (first entry)  
DE Excitatory amino acid receptor (EAA4a) fragment.  
KW Excitatory amino acid receptor; EAA4a; screening; detection;  
KW Identification; CNS; nervous disorders; antibody; oocyte; membrane  
OS protein: ion-channel; therapeutics.  
KW Homo sapiens.  
PN EP-578409-A.  
PD 12-JAN-1994.  
PF 24-JUN-1993; 304961.  
PR 24-JUN-1992; US-903456.  
PA (ELLI/) ELLIOTT C E.  
PA (KAMB/) KAMBOJ R.  
PA (NUTT/) NUTT S L.  
PI Kamboj R, Nutt SL, Elliott CE;  
DR WPI; 94-010248/02.  
PT Excitatory aminoacid for receptors with kainate binding activity  
PT - used to test ligands for CNS receptor interaction to identify  
PT cpds. useful against CNS disorders  
PS Disclosure: Figure 3a; 3pp: English.  
CC The excitatory amino acid receptor (See R45367) can be used to  
CC screen ligands. The ligands can be assayed for interaction by  
CC incubating them with cells, preferably oocytes, expressing the  
CC receptor or with membrane preparations from these cells, and  
CC assessing any interaction by determining receptor-ligand binding or  
CC ligand-mediated ion channel activation. This method is useful in  
CC the identification of therapeutics useful to treat CNS disorders in  
CC humans. Fragments of the receptor are useful for structural  
CC investigations and to raise antibodies. This fragment of the  
CC receptor (R45367) corresponds to amino acid residues 700 to 750.  
SQ Sequence 50 AA;

Query Match 60.5%; Score 52; DB 9; Length 50;

Best Local Similarity 60.0%; Pred. No. 3.71e+01;

Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 26 gglidskgyg 35

|||||:

QY 70 GALVDSKSYA 79

RESULT 2

ID W76238 standard; protein; 146 AA.

AC W76238:

DT 02-DEC-1998 (first entry)

DE Bacterial periplasmic binding protein fragment #13.

KW Protein-ligand binding pocket; PLBP; binding protein; ligand; modulator;

KW bacterial periplasmic binding protein; interaction energies; ischaemia;

KW basis set molecules; BSM; ionotropic glutamate receptors; treatment;

KW neuroprotectant; stroke; epilepsy; neuropathic pain.

OS Prokaryota.

PN W09838208-A2.

PD 03-SEP-1998.

PF 27-FEB-1998; U03951.

PR 28-FEB-1997; US-808804.

PA (BEAR-) BEARDSDEN BIO INC.

PI Sturgess M;

DR WPI; 98-495386/42.

PT Designing a model for a ligand-binding pocket in a protein and its

PT use for assessing ligand affinity - without the need to prepare test

PT ligand, for identifying selective antagonists for ionotropic

PT glutamate receptors, potentially useful as neuroprotectants with

PT reduced side effects

PS Disclosure: Page 215-216; 218pp; English.

CC W76236-W76250 are protein fragments used in the design of a model of a

CC protein-ligand binding pocket (PLBP) of a binding protein. A model is

CC constructed based on topographic similarity to the binding pocket of a

CC bacterial periplasmic binding protein. The model is refined by energy

CC minimisation, with a high affinity ligand in the binding pocket.

CC Interaction energies of basis set molecules (BSM) are obtained by

CC calculating energy of the models for the binding pocket BP and BSM

CC individually (Er and Em) and total energy (Erm) of bound complexes

CC formed. The model is used to predict the selectivity of a potential

CC ligand for a set of related PLBP's, specifically ionotropic glutamate

CC receptors. Identified ligands are potentially useful for studying

CC receptor binding and activity and as modulators of receptor activity

CC (or lead compounds for developing such compounds). The ligands are

CC potentially useful therapeutically, e.g. as neuroprotectants during

CC ischaemia and for treatment of stroke, head injuries, epilepsy,

CC neuropathic pain etc. The method allows affinity of ligands to be

CC estimated without having to prepare them and then test them, in vitro.

CC Ligands with high selectivity for particular glutamate receptors

CC should have fewer side effects than known receptor antagonists.

SQ Sequence 146 AA;

Query Match 60.5%; Score 52; DB 35; Length 146;

Best Local Similarity 60.0%; Pred. No. 3.71e+01;

Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 96 ggildskgyg 105

QY 70 GALVDSKSYA 79

RESULT 3

ID R63070 standard; Protein; 849 AA.

AC R63070:

DT 22-MAR-1995 (first entry)

DE Human EAA3d excitatory amino acid receptor.

KW excitatory amino acid receptor; human EAA3 receptor; kainate-type;

KW EAA3d receptor; central nervous system receptor; CNS.

OS Homo sapiens.

FH Key

FT peptide 1..30

FT protein 31..849

FT /label= signal\_peptide

FT /label= mature\_EAA3c\_receptor

FT /note= "truncated version of EAA3a in

FT which 40 amino acids have been eliminated

FT from the C-terminus and in which the last

FT 11 amino acids differ from those in the

FT corresponding region of EAA3a"

FT domain 31..562

FT /label= N-terminal\_domain

FT /note= "extracellular"

FT region 563..840

FT /label= hydrophobic\_region

FT domain 563..582

FT /label= TM-1

FT /note= "transmembrane domain"

FT domain 604..624

FT /label= TM-2

FT /note= "transmembrane domain"

FT domain 635..653

FT /label= TM-3

FT /note= "transmembrane domain"

FT domain 820..840

FT /label= TM-4

FT /note= "transmembrane domain"

FT domain 841..865

FT /label= C-terminal\_domain

FT /note= "extracellular"

PN CA2110933-A.

12-JUN-1994.

PF 08-DEC-1993; 110933.

PR 11-DEC-1992; US-989793.

PA (KAMB/) KAMBOJ R.

PA (ELLI/) ELLIOT C E.

PA (NUTT/) NUTT S L.

PI Elliot CE, Kamboj R, Nutt SL;

DR N-PSDB; Q81158.

DR N-PSDB; Q81158.

PT Polynucleotide encoding a human excitatory amino acid 3 receptor

PT or fragment - used to assay test ligands for their interaction

PT with a human CBS receptor.

PS Claim 13; Fig 1 and Fig 4C; 35pp; English.

CC The human EAA3a receptor (R60112) and its variants EAA3b, EAA3c and

CC EAA3d (R63068-R63070) occur naturally in human brain. They are all

CC members of the EAA3 receptor family, having extracellular N- and C-

CC termini and 4 internal hydrophobic domains which anchor the receptor

CC within the cell surface membrane. The receptors bind glutamate and

CC also exhibit binding properties characteristic of kainate-type EAA

CC receptors.

CC N.B. The cDNA and corresp. amino acid sequences for EAA3b, 3c and

CC 3d do not appear in full in the specification; the sequences have

CC been compiled by combining sequences in figures 1 and 4 according

CC to the description given in the disclosure.

SQ Sequence 849 AA;

Query Match 60.5%; Score 52; DB 11; Length 849;

Best Local Similarity 60.0%; Pred. No. 3.71e+01;

Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 700 ggildskgyg 709

QY 70 GALVDSKSYA 79

RESULT 4

ID R63069 standard; Protein; 865 AA.

AC R63069:

DT 22-MAR-1995 (first entry)

DE Human EAA3c excitatory amino acid receptor.

KW excitatory amino acid receptor; human EAA3 receptor; kainate-type;

KW EAA3c receptor; central nervous system receptor; CNS.

OS Homo sapiens.

FH Key

FT peptide 1..30

FT protein 31..865

FT /label= signal\_peptide

FT /label= mature\_EAA3c\_receptor

FT /note= "truncated version of EAA3a in

FT which 40 amino acids have been eliminated

FT from the C-terminus and in which the last

FT 11 amino acids differ from those in the

FT corresponding region of EAA3a"

FT domain 31..562

FT /label= N-terminal\_domain

FT /note= "extracellular"

FT region 563..840

FT /label= hydrophobic\_region

FT domain 563..582

FT /label= TM-1

FT /note= "transmembrane domain"

FT domain 604..624

FT /label= TM-2

FT /note= "transmembrane domain"

FT domain 635..653

FT /label= TM-3

FT /note= "transmembrane domain"

FT domain 820..840

FT /label= TM-4

FT /note= "transmembrane domain"

FT domain 841..865

FT /label= C-terminal\_domain

FT /note= "extracellular"

PN CA2110933-A.

PD 12-JUN-1994.  
 PF 08-DEC-1993; 110933.  
 PR 11-DEC-1992; US-989793.  
 PA (KAMB/) KAMBOJ R.  
 PA (ELLI/) ELLIOT C E.  
 PA (NUTT/) NUTT S L.  
 PI Elliot CE, Kamboj R, Nutt SL;  
 DR WPI: 94-255829/32.  
 DR N-PSDB: Q81157.  
 PT Polynucleotide encoding a human excitatory amino acid 3 receptor  
 PT or fragment - used to assay test ligands for their interaction  
 PT with a human CBS receptor.  
 PS Claim 13: Fig 1 and Fig 4B: 35pp; English.  
 CC The human EAA3a receptor (R60112) and its variants EAA3b, EAA3c and  
 CC EAA3d (R63068-R63070) occur naturally in human brain. They are all  
 CC members of the EAA3 receptor family, having extracellular N- and C-  
 CC termini and 4 internal hydrophobic domains which anchor the receptor  
 CC within the cell surface membrane. The receptors bind glutamate and  
 CC also exhibit binding properties characteristic of kainate-type EAA  
 CC receptors.  
 CC N.B. The cDNA and corresp. amino acid sequences for EAA3b, 3c and  
 CC 3d do not appear in full in the specification; the sequences have  
 CC been compiled by combining sequences in Figures 1 and 4 according  
 CC to the description given in the disclosure.  
 SQ Sequence 865 AA;  
 Query Match 60.5%; Score 52; DB 11; Length 865;  
 Best Local Similarity 60.0%; Pred. No. 3.71e+01;  
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Db 756 gglidskgyg 765  
 I:::||||:  
 Qy 70 GALVDSKSYA 79

RESULT 5  
 ID R11994 standard; Protein; 901 AA.  
 AC R11994;  
 DT 31-JUL-1991 (first entry)  
 DE Glutamate receptor 6.  
 KW Glutamate receptor 6; probe; ligand; drug screening.  
 OS Rattus rattus.  
 PN WO9105648-A.  
 PD 16-MAY-1991.  
 PF 25-OCT-1990; U06153.  
 PR 27-OCT-1989; US-428116.  
 PA (SALK ) SALK INST FOR BIOL STUD.  
 PI Heinemann SF, Boulter JR, Hollmann M, Bettler B, Jensen JE;  
 DR WPI: 91-164197/22.  
 DR N-PSDB: Q11854.  
 PT Glutamate receptors - used to screen for functional ligands and  
 PT identify and isolate further receptors  
 PS Disclosure; Fig 11; 109pp; English.  
 CC GluR6 has a Mr of 100,000.  
 CC The gene and protein can be used in drug screening, to  
 CC determine whether a substance is a functional ligand for the  
 CC receptor by monitoring ion channel activity.  
 CC See also Q11849-855.  
 SQ Sequence 901 AA;  
 Query Match 60.5%; Score 52; DB 2; Length 901;  
 Best Local Similarity 60.0%; Pred. No. 3.71e+01;  
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Db 756 gglidskgyg 765  
 I:::||||:  
 Qy 70 GALVDSKSYA 79

RESULT 6  
 ID R75883 standard; Protein; 905 AA.  
 AC R75883;  
 DT 10-JUL-1996 (first entry)

DE Human EAA3 receptor (Q-591).  
 KW Glutamate receptor; EAA3 receptor; excitatory amino acid;  
 OS CNS receptor; RNA editing.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT peptide 1..30  
 FT /label= Sig\_peptide  
 FT 31..905  
 FT /label= Mat\_protein  
 PN WO9517508-A2.  
 PD 29-JUN-1995.  
 PD 21-DEC-1994; CA0705.  
 PR 23-DEC-1993; US-172188.  
 PA (ALLX ) ALLELIX BIOPHARMACEUTICALS INC.  
 PI Kamboj R, Nutt S;  
 DR WPI: 95-240670/31.  
 DR N-PSDB: Q91231.  
 PT Identification of human CNS receptor ligand - and identification of  
 PT agents that modulate editing of human CNS receptors  
 PS Claim 4; Fig 5; 59pp; English.  
 CC The amino acid sequence (R75883) of human EAA3 receptor deduced from  
 CC a genomic DNA sequence (Q91231) differed from that deduced from  
 CC EAA3 cDNA at position 591 of the mature protein; the genomic sequence  
 CC coded for glutamine at this position, and the cDNA sequence for  
 CC arginine. This was due to a single nucleotide change of A to G at  
 CC position 2279 as a result of RNA editing. The edited and unedited  
 CC receptors were functionally distinct and appeared to be selectively  
 CC expressed. CNS receptors subject to such editing (see also R75882  
 CC and R75884) have value in screening cpds. for potential therapeutic  
 CC utility.  
 SQ Sequence 905 AA;  
 Query Match 60.5%; Score 52; DB 16; Length 905;  
 Best Local Similarity 60.0%; Pred. No. 3.71e+01;  
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Db 756 gglidskgyg 765  
 I:::||||:  
 Qy 70 GALVDSKSYA 79

RESULT 7  
 ID R63068 standard; Protein; 905 AA.  
 AC R63068;  
 DT 22-MAR-1995 (first entry)  
 DE Human EAA3b excitatory amino acid receptor.  
 KW excitatory amino acid receptor; human EAA3 receptor; kainate-type;  
 KW EAA3a receptor; central nervous system receptor; CNS.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT peptide 1..30  
 FT /label= signal\_peptide  
 FT protein 31..905  
 FT /label= mature\_EAA3b\_receptor  
 FT /note= "differs from EAA3a by one amino acid,  
 FT i.e. Asp at position +639 of mature  
 FT EAA3a is replaced by Asn in EAA3b"  
 FT 31..562  
 FT /label= N-terminal\_domain  
 FT /note= "extracellular"  
 FT 563..840  
 FT /label= hydrophobic\_region  
 FT 563..582  
 FT /label= TM-1  
 FT /note= "transmembrane domain"  
 FT 604..624  
 FT /label= TM-2  
 FT /note= "transmembrane domain"  
 FT 635..653  
 FT /label= TM-3  
 FT /note= "transmembrane domain"  
 FT 820..840  
 FT /label= TM-4

```

FT FT domain /note= "transmembrane domain"
FT FT 841...905
FT FT /label= C-terminal_domain
FT FT /note= "extracellular"
PN CA2110933-A.
PD 12-JUN-1994.
PF 08-DEC-1993; 110933.
PR 11-DEC-1992; US-989793.
PA (KAMB/) KAMBOJ R.
PA (ELLI/) ELLIOT C E.
PA (NUTT/) NUTT S L.
PI Elliot CE, Kamboj R, Nutt SL;
DR WPI: 94-255829/32.
DR N-PSDB: Q70097.
PT Polynucleotide encoding a human excitatory amino acid 3 receptor
PT or fragment - used to assay test ligands for their interaction
PT with a human CBS receptor.
PS Claim 13; Fig 1; 35pp; English.
CC The human EAA3a receptor (R60112) and its variants EAA3b, EAA3c and
CC EAA3d (R63068-R63070) occur naturally in human brain. They are all
CC members of the EAA3 receptor family, having extracellular N- and C-
CC termini and 4 internal hydrophobic domains which anchor the receptor
CC within the cell surface membrane. The receptors bind glutamate and
CC also exhibit binding properties characteristic of kainate-type EAA
CC receptors.
CC N.B. The cDNA and corresp. amino acid sequences for EAA3b, 3c and
CC 3d do not appear in full in the specification; the sequences have
CC been compiled by combining sequences in Figures 1 and 4 according
CC to the description given in the disclosure.
CC Sequence 905 AA;
SQ
Query Match 60.5%; Score 52; DB 11; Length 905;
Best Local Similarity 60.0%; Pred. No. 3.71e+01;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 756 gglidskgyg 765
QY 70 GALVDSKSYA 79
RESULT 8
ID R60112 standard; Protein: 905 AA.
AC R60112.
DT 22-MAR-1995 (first entry)
DE Human EAA3a excitatory amino acid receptor.
KW excitatory amino acid receptor; human EAA3 receptor; kainate-type;
KW EAA3a receptor.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= signal_peptide
FT protein 31..905
FT /label= mature_EAA3a_receptor
FT domain 31..562
FT /label= N-terminal_domain
FT /note= "extracellular"
FT region 563..840
FT /label= hydrophobic_region
FT domain 563..582
FT /label= TM-1
FT /note= "transmembrane domain"
FT domain 604..624
FT /label= TM-2
FT /note= "transmembrane domain"
FT domain 635..653
FT /label= TM-3
FT /note= "transmembrane domain"
FT domain 820..840
FT /label= TM-4
FT /note= "transmembrane domain"
FT domain 841..905
FT /label= C-terminal_domain
FT /note= "extracellular"

```

```

PN CA2110933-A.
PD 12-JUN-1994.
PF 08-DEC-1993; 110933.
PR 11-DEC-1992; US-989793.
PA (KAMB/) KAMBOJ R.
PA (ELLI/) ELLIOT C E.
PA (NUTT/) NUTT S L.
PI Elliot CE, Kamboj R, Nutt SL;
DR WPI: 94-255829/32.
DR N-PSDB: Q70097.
PT Polynucleotide encoding a human excitatory amino acid 3 receptor
PT or fragment - used to assay test ligands for their interaction
PT with a human CBS receptor.
PS Claim 13; Fig 1; 35pp; English.
CC The sequence coding for the EAA3a receptor was isolated by probing a
CC human foetal brain cDNA library. The EAA3a receptor and naturally
CC occurring variants of it (i.e. EAA3b, EAA3c and EAA3d) are members
CC of the EAA3 receptor family; they all have extracellular N- and C-
CC termini and 4 internal hydrophobic domains which anchor the receptor
CC within the cell surface membrane. The receptors bind glutamate and
CC also exhibit binding properties characteristic of kainate-type EAA
CC receptors.
CC Sequence 905 AA;
SQ
Query Match 60.5%; Score 52; DB 11; Length 905;
Best Local Similarity 60.0%; Pred. No. 3.71e+01;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 756 gglidskgyg 765
QY 70 GALVDSKSYA 79
RESULT 9
ID R75884 standard; Protein: 908 AA.
AC R75884;
DT 10-JUL-1996 (first entry)
DE Human EAA4 receptor (I532, Y536, Q586).
KW Glutamate receptor; EAA4 receptor; excitatory amino acid;
KW CNS receptor; RNA editing.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..31
FT /label= sig_peptide
FT protein 32..908
FT /label= Mat_protein
FT WO9517508-A2.
PN 29-JUN-1995.
PF 21-DEC-1994; CA0705.
PR 23-DEC-1993; US-172188.
PA (ALIX ) ALLELIX BIOPHARMACEUTICALS INC.
PA Kamboj R, Nutt S;
DR WPI: 95-240670/31.
DR N-PSDB: Q91232.
PT Identification of human CNS receptor ligand - and identification of
PT agents that modulate editing of human CNS receptors
PS Claim 5; Fig 6; 59pp; English.
CC The amino acid sequence (R75884) of human EAA4 receptor deduced from
CC a genomic DNA sequence (Q91232) differed from that deduced from
CC EAA4 cDNA; the genomic sequence coded for Ile, Tyr and Gln at
CC positions 532, 536 and 586, respectively, in the mature protein,
CC and the cDNA sequence for Val, Cys and Arg at these positions.
CC These changes were each a result of A to G substitutions in the
CC coding sequence and a consequence of RNA editing. The edited and
CC unedited receptors were functionally distinct and appeared to be
CC selectively expressed. CNS receptors subject to such editing (see
CC also R75882 and R75884) have value in screening cpds. for potential
CC therapeutic utility.
CC Sequence 908 AA;
SQ
Query Match 60.5%; Score 52; DB 16; Length 908;
Best Local Similarity 60.0%; Pred. No. 3.71e+01;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

Db 756 gglidskgyg 765  
 QY 70 GALVDSKSYA 79

RESULT 10  
 ID R45367 standard; Protein: 908 AA.  
 AC R45367;  
 DT 07-JUL-1994 (first entry)  
 DE Excitatory amino acid receptor (EAA4a receptor).  
 KW Excitatory amino acid receptor; EAA4a; screening; detection;  
 KW identification; CNS; nervous disorders; antibody; oocyte; membrane  
 KW protein; ion-channel; therapeutics.  
 OS Homo sapiens.  
 PN EP-578409-A.  
 PD 12-JAN-1994.  
 PR 24-JUN-1993: 304961.  
 PR 24-JUN-1992; US-903456.  
 PA (ELLI/) ELLIOTT C E.  
 PA (KAMB/) KAMBOJ R.  
 PA (NUTT/) NUTT S L.  
 PI Kamboj R, Nutt SL, Elliott CE;  
 DR WPI: 94-010248/02.  
 DR N-PSDB; Q54482.  
 PT Excitatory amino acid for receptors with kainate binding activity  
 PT - used to test ligands for CNS receptor interaction to identify  
 PT cpds. useful against CNS disorders  
 PS Claim 13; Figure 1; 37pp; English.  
 CC The excitatory amino acid receptor can be used to screen ligands.  
 CC The ligands can be assayed for interaction by incubating them with  
 CC cells, preferably oocytes, expressing the receptor or with membrane  
 CC preparations from these cells, and assessing any interaction by  
 CC determining receptor-ligand binding or ligand-mediated ion channel  
 CC activation. This method is useful in the identification of  
 CC therapeutics useful to treat CNS disorders in humans. Fragments of  
 CC the receptor are useful for structural investigations and to raise  
 CC antibodies.  
 SQ Sequence 908 AA;

Query Match 60.5%; Score 52; DB 9; Length 908;  
 Best Local Similarity 60.0%; Pred. No. 3.71e+01;  
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 756 gglidskgyg 765  
 QY 70 GALVDSKSYA 79

RESULT 11  
 ID R50179 standard; Protein: 919 AA.  
 AC R50179;  
 DT 17-OCT-1994 (first entry)  
 DE Excitatory amino acid receptor (EAA5).  
 KW Excitatory amino acid receptor; EAA; identification; detection;  
 KW CNS; central nervous system; therapeutic; antibody; ligand;  
 KW screening.  
 OS Homo sapiens.  
 PN EP-588642-A.  
 PD 23-MAR-1994.  
 PR 16-SEP-1993: 307325.  
 PR 17-SEP-1992; US-945210.  
 PA (ELLI/) ELLIOTT C E.  
 PA (KAMB/) KAMBOJ R.  
 PA (NUTT/) NUTT S L.  
 PI Elliott CE, Kamboj R, Nutt SL;  
 DR WPI: 94-094202/12.  
 DR N-PSDB; Q44685.  
 PT Nucleic acid encoding human excitatory amino acid receptors - used  
 PT for producing receptors and in assays for test ligands for  
 PT binding to human CNS receptors  
 PS Claim 12; page 16-19; 34pp; English.  
 CC Nucleotides which encode the excitatory amino acid receptors (EAA's)

CC can be used for the production of the receptors, to identify  
 CC sequence related genes or for locating the receptor encoding DNA in  
 CC a specimen. Recombinant cells which produce the receptors, or their  
 CC membrane preparations, can be used for assaying a test ligand for  
 CC binding to a human CNS receptor to develop therapeutics. The  
 CC receptors can themselves be used in the production of antibodies for  
 CC use in detection methods.  
 SQ Sequence 919 AA;

Query Match 60.5%; Score 52; DB 10; Length 919;  
 Best Local Similarity 60.0%; Pred. No. 3.71e+01;  
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 757 gglidskgyg 766  
 QY 70 GALVDSKSYA 79

RESULT 12  
 ID R11993 standard; Protein: 920 AA.  
 AC R11993;  
 DT 31-JUL-1991 (first entry)  
 DE Glutamate receptor 5-1.  
 KW Glutamate receptor 5-1; probe; ligand; drug screening.  
 OS Rattus rattus.  
 FH Key Location/Qualifiers  
 FT peptide 1..30  
 FT /label= sig\_peptide  
 FT protein 31..920  
 FT /label= mat\_protein  
 FT region 402..416  
 FT /label= insertion  
 PN W09106648-A.  
 PD 16-MAY-1991.  
 PR 25-OCT-1989; U06153.  
 PR 27-OCT-1989; US-428116.  
 PA (SALK ) SALK INST FOR BIOL STUD.  
 PI Heinemann SF, Boulter JR, Hollmann M, Bettler B, Jensen JE;  
 DR WPI: 91-164197/22.  
 DR N-PSDB; Q11853.  
 PT Glutamate receptors - used to screen for functional ligands and  
 PT identify and isolate further receptors  
 PS Disclosure; Fig 10; 109pp; English.  
 CC GluR5-1 has a 15 amino acid insert (see features) compared to the  
 CC shorter variant GluR5-2 and is unique among the receptors GluR1-7.  
 CC It has a Mr of 100,000. The signal sequence cleavage site is after  
 CC a Pro, which is atypical.  
 CC The gene and protein can be used in drug screening, to  
 CC determine whether a substance is a functional ligand for the  
 CC receptor by monitoring ion channel activity.  
 CC See also Q11849-855.  
 SQ Sequence 920 AA;

Query Match 60.5%; Score 52; DB 2; Length 920;  
 Best Local Similarity 60.0%; Pred. No. 3.71e+01;  
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 771 gglidskgyg 780  
 QY 70 GALVDSKSYA 79

RESULT 13  
 ID R45369 standard; Protein: 50 AA.  
 AC R45369;  
 DT 07-JUL-1994 (first entry)  
 DE Excitatory amino acid receptor (EAA4b) fragment.  
 KW Excitatory amino acid receptor; EAA4a; screening; detection;  
 KW identification; CNS; nervous disorders; antibody; oocyte; membrane  
 KW protein; ion-channel; therapeutics.  
 OS Homo sapiens.  
 PN EP-578409-A.  
 PD 12-JAN-1994.

PF 24-JUN-1993; 304961.  
 PR 24-JUN-1992; US-903456.  
 PA (ELLI/) ELLIOTT C E.  
 PA (KAMB/) KAMBOJ R.  
 PA (NUTT/) NUTT S L.  
 PI Kamboj R, Nutt SL, Elliott CE;  
 DR WPI: 94-010248/02.  
 PT Excitatory aminoacid for receptors with kainate binding activity  
 PT - used to test ligands for CNS receptor interaction to identify  
 PT cpds. useful against CNS disorders  
 PS Disclosure: Figure 3a; 37pp; English.  
 CC The excitatory amino acid receptor EAA4a can be used to screen  
 CC ligands. The ligands can be assayed for interaction by incubating  
 CC them with cells, preferably oocytes, expressing the receptor or with  
 CC membrane preparations from these cells, and assessing any  
 CC in-eraction by determining receptor-ligand binding or  
 CC ligand-mediated ion channel activation. This method is useful in  
 CC the identification of therapeutics useful to treat CNS disorders in  
 CC humans. Fragments of the receptor are useful for structural  
 CC investigations and to raise antibodies. This fragment of the  
 CC receptor EAA4b corresponds to amino acid residues 700 to 750 and has  
 CC aspartic acid at position 727 as opposed to glycine which occurs in  
 CC the corresponding position in receptor-EAA4a.  
 SQ Sequence 50 AA;

Query Match 59.3%; Score 51; DB 9; Length 50;  
 Best Local Similarity 60.0%; Pred. No. 4.90e+01;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 26 gdlidskgvg 35  
 | :|:|:|:  
 QY 70 GALVDSKSYA 79

## RESULT 14

ID P60011 standard; Protein; 259 AA.  
 AC P60011;  
 DT 30-JUL-1991 (first entry)  
 DE Sequence of oncogene v-fms.  
 KW Therapy: diagnosis; insulin.  
 PN EP-192392-A.  
 PD 27-AUG-1986.  
 PF 10-FEB-1986; 300894.  
 PR 12-FEB-1985; US-700776.  
 PA (GETH ) GENENTECH INC.  
 PI Bell JR, Ulrich A, Ramachandran J;  
 DR WPI: 86-226966/35.  
 PT New DNA encoding insulin receptor or its fragments - used for  
 PT synthesis of receptor and mutants for therapeutic and diagnostic  
 PT use  
 PS Example: Fig 5; 62pp; English.  
 CC A mutant IR is claimed which may have a mutated alpha-chain, esp. at  
 CC the precursor processing site. The beta-chain may be mutated, e.g.  
 CC by deletion of the transmembrane sequence; the tyrosine kinase  
 CC activity may be inactivated. Fig. 5 is a comparison of oncogene and  
 CC human EGF receptor sequences with that of HIR in the cytoplasmic  
 CC domain of the insulin receptor beta subunit.  
 SQ Sequence 259 AA;

Query Match 58.1%; Score 50; DB 3; Length 259;  
 Best Local Similarity 53.8%; Pred. No. 6.45e+01;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 191 npypgilvnskfy 203  
 : | | | | |  
 QY 66 DPSTGALVDSKSY 78

## RESULT 15

ID W41513 standard; Protein; 371 AA.  
 AC W41513;  
 DT 17-JUN-1998 (first entry)  
 DE N. meningitidis alpha-2,3-sialyltransferase protein.

KW Alpha-2,3-sialyltransferase; sialic acid; acceptor; lipid;  
 KW biologically active oligosaccharide; sialyl-modified protein.  
 OS Neisseria meningitidis.  
 PN WO9747749-A1.  
 PD 18-DEC-1997.

PF 10-JUN-1997; CA03990.  
 PR 06-JUN-1997; US-872485.  
 PR 10-JUN-1996; US-019520.  
 PA (CANA ) NAT RES COUNCIL CANADA.  
 PI Gilbert M, Jennings MP, Wakarchuk WW, Young NW;  
 DR WPI: 98-052313/05.  
 DR N-PSDB; V04125.

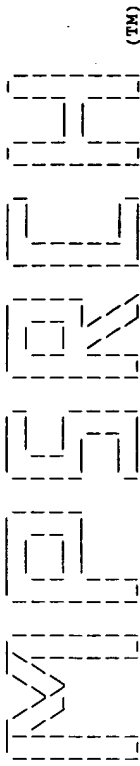
PT Nucleic acid sequence encoding Neisseria  
 PT alpha-2,3-sialyltransferase - useful to add sialic acid to acceptor  
 PT with terminal galactose residue for synthesis of biologically active  
 PT oligosaccharide.  
 PS Claim 31; Page 38; 50pp; English.  
 CC This sequence represents an alpha-2,3-sialyltransferase which has been  
 CC isolated from Neisseria meningitidis. The protein can be used as a  
 CC reagent for adding a sialic acid residue to an acceptor having a  
 CC terminal galactose residue, e.g. in synthesis of biologically active  
 CC oligosaccharides or sialyl-modified proteins or lipids. The nucleic  
 CC acid sequence can be used for the recombinant production of  
 CC alpha-2,3-sialyltransferase.  
 SQ Sequence 371 AA;

Query Match 58.1%; Score 50; DB 29; Length 371;  
 Best Local Similarity 53.8%; Pred. No. 6.45e+01;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 164 ddgtgnliqssy 176  
 | :| | :|:  
 QY 66 DPSTGALVDSKSY 78

Search completed: Wed Sep 1 16:30:56 1999  
 Job time : 25 secs.

\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:30:02 1999; MasPar time 3.68 Seconds  
Tabular output not generated. 152.280 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (66-79) from PCTUS9913024.ppt (11 of 12)  
Perfect Score: 86  
Sequence: 1 DPSTGALVDSKSYA 14

Scoring table: PAM 150  
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 24.770; Variance 28.525; scale 0.868

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	52	60.5	864	2 A43954	glutamate receptor be	4.06e+00
2	52	60.5	869	2 S35792	glutamate receptor G1	4.06e+00
3	52	60.5	908	2 A54260	glutamate receptor 6	4.06e+00
4	52	60.5	908	2 S19098	glutamate receptor G1	4.06e+00
5	52	60.5	918	2 I58178	glutamate receptor -	4.06e+00
6	52	60.5	919	2 S19810	glutamate receptor G1	4.06e+00
7	52	60.5	919	2 I53474	kainate receptor - ra	4.06e+00
8	52	60.5	949	2 S19808	glutamate receptor G1	4.06e+00
9	51	59.3	469	2 S17813	crtJ protein - rhodob	6.50e+00
10	51	58.3	563	2 S04886	mutL protein - Vibrio	6.50e+00
11	50	58.1	331	2 A28063	glutamin-(asparagin-)	1.06e+01
12	50	58.1	542	2 J55507	monocarboxylate trans	1.06e+01
13	50	58.1	941	1 TVNMVD	protein-tyrosine kina	1.06e+01
14	50	58.1	972	1 TVHUMD	macrophage colony-sti	1.06e+01
15	50	58.1	980	1 TVCTMD	macrophage colony-sti	1.06e+01
16	50	58.1	1133	2 T01757	hypothetical protein	1.70e+01
17	49	57.0	235	2 S48382	hypothetical protein	1.70e+01
18	49	57.0	319	2 A69756	adhesion protein homo	1.70e+01
19	49	57.0	357	2 A54512	serine-repeat antigen	1.70e+01
20	49	57.0	427	2 A34639	parasitophorous vacuo	1.70e+01
21	49	57.0	477	1 VGBEGI	glycoprotein E - suid	1.70e+01
22	49	57.0	635	1 SCCHAL	S-aminolevulinatase syn	1.70e+01
23	49	57.0	966	2 G69189	hypothetical protein	1.70e+01

24	49	57.0	989	2 A54505	serine-repeat antigen	1.70e+01
25	49	57.0	997	2 B71617	SERA antigen/papain-1	1.70e+01
26	48	55.8	195	2 E70555	hypothetical protein	2.70e+01
27	48	55.8	427	2 S11889	S-locus-specific glyco	2.70e+01
28	48	55.8	432	2 B53223	SLR2 protein (S6 alle	2.70e+01
29	48	55.8	439	2 A53223	SLR2 protein (S2 alle	2.70e+01
30	48	55.8	448	2 P50047	serine proteinase (EC	2.70e+01
31	48	55.8	676	2 S69783	outer membrane protei	2.70e+01
32	48	55.8	683	2 S69780	outer membrane protei	2.70e+01
33	48	55.8	707	2 S69781	outer membrane protei	2.70e+01
34	48	55.8	712	2 S69782	outer membrane protei	2.70e+01
35	47	54.7	232	2 A71068	hypothetical protein	4.25e+01
36	47	54.7	257	2 F70166	hypothetical protein	4.25e+01
37	47	54.7	398	2 B71284	probable periplasmic	4.25e+01
38	47	54.7	432	2 S67821	GumE protein - xantho	4.25e+01
39	47	54.7	434	2 S04534	invasin precursor - y	4.25e+01
40	47	54.7	501	2 S53583	splicing factor SF3a6	4.25e+01
41	47	54.7	501	2 A55749	spliceosome-associate	4.25e+01
42	47	54.7	503	1 YF8YAC	phenylalanine-tRNA l	4.25e+01
43	47	54.7	606	2 A70960	probable pckA protein	4.25e+01
44	47	54.7	747	2 S37694	gene PC326 protein -	4.25e+01
45	47	54.7	2233	1 ZLNZP3	genome polyprotein -	4.25e+01

ALIGNMENTS

RESULT 1  
ENTRY A43954 #type complete  
TITLE glutamate receptor beta-2 chain - mouse  
ALTERNATE\_NAMES kainate receptor  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 06-Sep-1996  
ACCESSIONS A43954; I48131  
REFERENCE A43954  
#authors Morita, T.; Sakimura, K.; Kushiya, E.; Yamazaki, M.; Meguro, H.; Araki, K.; Abe, T.; Mori, K.J.; Mishina, M.  
#journal Brain Res. Mol. Brain Res. (1992) 14:143-146  
#title Cloning and functional expression of a cDNA encoding the mouse beta 2 subunit of the kainate-selective glutamate receptor channel.  
#cross-references MUID:92356790  
#accession A43954  
#status nucleic acid sequence not shown  
#molecule\_type mRNA  
#residues 1-864 #label MOR  
#cross-references GB:D10054; NID:g220408; PID:g220409  
#experimental\_source forebrain  
#note #sequence extracted from NCBI backbone (NCBIP:110532)

GENETICS  
#gene pGluRbeta-2  
CLASSIFICATION #superfamily glutamate receptor; glutamate receptor homology  
KEYWORDS neurotransmitter receptor; transmembrane protein  
FEATURE 435-855 #domain glutamate receptor homology #label GRH  
SUMMARY #length 864 #molecular-weight 97353 #checksum 9801  
Query Match 60.5%; Score 52; DB 2; Length 864;  
Best Local Similarity 60.0%; Pred. No. 4.06e+00;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
Db 756 GGLIDSKGY 765  
QY 70 GALVDSKSYA 79  
RESULT 2  
ENTRY S35792 #type complete  
TITLE glutamate receptor GluR6C - mouse  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 08-Sep-1997  
ACCESSIONS S35792



```
REFERENCE S35792
#authors Gregor, P.
#submission submitted to the EMBL Data Library, June 1992
#accession S35792
#status preliminary
#molecule_type mRNA
##residues 1-869 ##label GRE
##cross-references EMBL:X56117; NID:g312493; PID:g312494
CLASSIFICATION #superfamily glutamate receptor; glutamate receptor homology
KEYWORDS neurotransmitter receptor; transmembrane protein
FEATURE
435-855 #domain glutamate receptor homology #label GRH
SUMMARY #length 869 #molecular-weight 97807 #checksum 1988

Query Match 60.5%; Score 52; DB 2; Length 869;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 756 GGLIDSKGYG 765
QY 70 GALVDSKSYA 79

RESULT 3
ENTRY A54260 #type complete
TITLE glutamate receptor 6 kainate-preferring precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change
29-Aug-1997
ACCESSIONS A54260
REFERENCE A54260
#authors Paschen, W.; Blackstone, C.D.; Huganir, R.L.; Ross, C.A.
#journal Genomics (1994) 20:435-440
#title Human GluR6 kainate receptor (GRIK2): molecular cloning,
expression, polymorphism, and chromosomal assignment.
#cross-references MVID:94307730
#accession A54260
##status preliminary; not compared with conceptual translation
##molecule_type mRNA
##residues 1-908 ##label PAS
GENETICS
#gene GDB:GRIK2
##cross-references GDB:l31461; OMIM:138244
#map_position 6q16.3-q621
CLASSIFICATION #superfamily glutamate receptor; glutamate receptor homology
KEYWORDS neurotransmitter receptor
FEATURE
435-855 #domain glutamate receptor homology #label GRH
SUMMARY #length 908 #molecular-weight 102612 #checksum 2705

Query Match 60.5%; Score 52; DB 2; Length 908;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 756 GGLIDSKGYG 765
QY 70 GALVDSKSYA 79

RESULT 4
ENTRY S19098 #type complete
TITLE glutamate receptor GluR6 precursor - rat
ALTERNATE_NAMES kainate receptor
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change
08-Sep-1997
ACCESSIONS S19098; S16234; S19809
REFERENCE S19098
#authors Bettler, B.
#submission submitted to the EMBL Data Library, December 1991
#accession S19098
#molecule_type mRNA
##residues 1-908 ##label BET
```

```
##cross-references EMBL:Z11548; NID:g56281; PID:g56282
REFERENCE S16234
#authors Egebjerg, J.; Bettler, B.; Hermans-Borgmeyer, I.; Heinemann,
S.
#journal Nature (1991) 351:745-748
#title Cloning of a cDNA for a glutamate receptor subunit activated
by kainate but not AMPA.
#cross-references MVID:91287799
#accession S16234
#status nucleic acid sequence not shown
#molecule_type mRNA
##residues 1-880,'POLL' ##label EGE
##cross-references EMBL:Z11548
REFERENCE S19806
#authors Seeburg, P.H.
#submission submitted to the EMBL Data Library, February 1992
#accession S19809
#molecule_type mRNA
##residues 1-620,'Q',622-908 ##label SEE
##cross-references EMBL:Z11715; NID:g56279; PID:g56280
CLASSIFICATION #superfamily glutamate receptor; glutamate receptor homology
KEYWORDS neurotransmitter receptor; transmembrane protein
FEATURE
1-31 #domain signal sequence #status predicted #label SIG\
32-908 #product glutamate receptor GluR6 #status predicted
#label MAR\
435-855 #domain glutamate receptor homology #label GRH\
505-521 #domain transmembrane #status predicted #label TM1\
563-581 #domain transmembrane #status predicted #label TM2\
635-653 #domain transmembrane #status predicted #label TM3\
820-840 #domain transmembrane #status predicted #label TM4
SUMMARY #length 908 #molecular-weight 102423 #checksum 3737

Query Match 60.5%; Score 52; DB 2; Length 908;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 756 GGLIDSKGYG 765
QY 70 GALVDSKSYA 79

RESULT 5
ENTRY I58178 #type complete
TITLE glutamate receptor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
28-Aug-1998
ACCESSIONS I58178
REFERENCE I58178
#authors Gregor, P.; O'Hara, B.F.; Yang, X.; Uhl, G.R.
#journal Neuroreport (1993) 4:1343-1346
#title Expression and novel subunit isoforms of glutamate receptor
genes GluR5 and GluR6.
#cross-references MVID:94083547
#accession I58178
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
##residues 1-918 ##label RES
##cross-references GB:L19058; NID:g455447; PID:g455448
GENETICS
#gene GDB:GRIK1; GLUR5
##cross-references GDB:l31462; OMIM:138245
#map_position 21q22.1-21q22.2
CLASSIFICATION #superfamily glutamate receptor; glutamate receptor homology
KEYWORDS neurotransmitter receptor
SUMMARY #length 918 #molecular-weight 103980 #checksum 1748

Query Match 60.5%; Score 52; DB 2; Length 918;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 771 GGLIDSKGYG 780
```

```

Qy 70 GALVDSKSYA 79
      1:|||||:
      6
RESULT 6
ENTRY 6
TITLE 6
ALTERNATE_NAMES 6
ORGANISM 6
DATE 6
ACCESSIONS 6
REFERENCE 6
AUTHORS 6
SUBMISSION 6
#cross-references EMBL:211716
#molecule_type mRNA
#residues 1-919 ##label SEE
#cross-references EMBL:211716
REFERENCE 6
AUTHORS 6
#journal 6
#title 6
#cross-references MUID:92153417
#accession 6
#molecule_type mRNA
#residues 32-117,'T',119-175,'T',177-209,'D',211,'DD',214-289,'P',
291-340,'P',342-354,'A',356-459,'DR',462-466,'ID',
469-919 ##label BET
##cross-references GB:M83552; NID:G204391; PID:G204392
##experimental_source cerebellum
COMMENT 6
Glutamate receptors form part of a major excitatory
neurotransmitter system of the brain and play roles in synapse
formation, in learning and memory, and in the pathogenesis of
neurodegenerative diseases.
CLASSIFICATION 6
KEYWORDS 6
transmembrane protein
FEATURE 6
438-856 #domain glutamate receptor homology #label GRH\
507-524 #domain transmembrane #status predicted #label TM1\
565-584 #domain transmembrane #status predicted #label TM2\
608-626 #domain transmembrane #status predicted #label TM3\
637-663 #domain transmembrane #status predicted #label TM4\
821-841 #domain transmembrane #status predicted #label TM5\
415 #binding_site carbohydrate (Ash) (covalent) #status
predicted\
668 #binding_site phosphate (Ser) (covalent) #status
predicted
SUMMARY 6
#length 919 #molecular-weight 103991 #checksum 9539
Query Match 60.5%; Score 52; DB 2; Length 919;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 757 GGLIDSKGYG 766
      1:|||||:
      70 GALVDSKSYA 79
      1:|||||:
      7
RESULT 7
ENTRY 7
TITLE 7
ALTERNATE_NAMES 7
ORGANISM 7
DATE 7
ACCESSIONS 7
REFERENCE 7
AUTHORS 7
SUBMISSION 7
#cross-references EMBL:211716
#molecule_type mRNA
#residues 1-919 ##label SEE
#cross-references EMBL:211716
REFERENCE 7
AUTHORS 7
#journal 7
#title 7
#cross-references MUID:92153417
#accession 7
#molecule_type mRNA
#residues 32-117,'T',119-175,'T',177-209,'D',211,'DD',214-289,'P',
291-340,'P',342-354,'A',356-459,'DR',462-466,'ID',
469-919 ##label BET
##cross-references GB:M83552; NID:G204391; PID:G204392
##experimental_source cerebellum
COMMENT 7
Glutamate receptors form part of a major excitatory
neurotransmitter system of the brain and play roles in synapse
formation, in learning and memory, and in the pathogenesis of
neurodegenerative diseases.
CLASSIFICATION 7
KEYWORDS 7
transmembrane protein
FEATURE 7
438-856 #domain glutamate receptor homology #label GRH\
507-524 #domain transmembrane #status predicted #label TM1\
565-584 #domain transmembrane #status predicted #label TM2\
608-626 #domain transmembrane #status predicted #label TM3\
637-663 #domain transmembrane #status predicted #label TM4\
821-841 #domain transmembrane #status predicted #label TM5\
415 #binding_site carbohydrate (Ash) (covalent) #status
predicted\
668 #binding_site phosphate (Ser) (covalent) #status
predicted
SUMMARY 7
#length 919 #molecular-weight 103991 #checksum 9539
Query Match 60.5%; Score 52; DB 2; Length 919;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 757 GGLIDSKGYG 766
      1:|||||:
      70 GALVDSKSYA 79
      1:|||||:
      7
RESULT 7
ENTRY 7
TITLE 7
ALTERNATE_NAMES 7
ORGANISM 7
DATE 7
ACCESSIONS 7
REFERENCE 7
AUTHORS 7
SUBMISSION 7
#cross-references EMBL:211716
#molecule_type mRNA
#residues 1-919 ##label SEE
#cross-references EMBL:211716
REFERENCE 7
AUTHORS 7
#journal 7
#title 7
#cross-references MUID:92153417
#accession 7
#molecule_type mRNA
#residues 32-117,'T',119-175,'T',177-209,'D',211,'DD',214-289,'P',
291-340,'P',342-354,'A',356-459,'DR',462-466,'ID',
469-919 ##label BET
##cross-references GB:M83552; NID:G204391; PID:G204392
##experimental_source cerebellum
COMMENT 7
Glutamate receptors form part of a major excitatory
neurotransmitter system of the brain and play roles in synapse
formation, in learning and memory, and in the pathogenesis of
neurodegenerative diseases.
CLASSIFICATION 7
KEYWORDS 7
transmembrane protein
FEATURE 7
438-856 #domain glutamate receptor homology #label GRH\
507-524 #domain transmembrane #status predicted #label TM1\
565-584 #domain transmembrane #status predicted #label TM2\
608-626 #domain transmembrane #status predicted #label TM3\
637-663 #domain transmembrane #status predicted #label TM4\
821-841 #domain transmembrane #status predicted #label TM5\
415 #binding_site carbohydrate (Ash) (covalent) #status
predicted\
668 #binding_site phosphate (Ser) (covalent) #status
predicted
SUMMARY 7
#length 919 #molecular-weight 103991 #checksum 9539
Query Match 60.5%; Score 52; DB 2; Length 919;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 757 GGLIDSKGYG 766
      1:|||||:
      70 GALVDSKSYA 79
      1:|||||:
      7
RESULT 7
ENTRY 7
TITLE 7
ALTERNATE_NAMES 7
ORGANISM 7
DATE 7
ACCESSIONS 7
REFERENCE 7
AUTHORS 7
SUBMISSION 7
#cross-references EMBL:211716
#molecule_type mRNA
#residues 1-919 ##label SEE
#cross-references EMBL:211716
REFERENCE 7
AUTHORS 7
#journal 7
#title 7
#cross-references MUID:92153417
#accession 7
#molecule_type mRNA
#residues 32-117,'T',119-175,'T',177-209,'D',211,'DD',214-289,'P',
291-340,'P',342-354,'A',356-459,'DR',462-466,'ID',
469-919 ##label BET
##cross-references GB:M83552; NID:G204391; PID:G204392
##experimental_source cerebellum
COMMENT 7
Glutamate receptors form part of a major excitatory
neurotransmitter system of the brain and play roles in synapse
formation, in learning and memory, and in the pathogenesis of
neurodegenerative diseases.
CLASSIFICATION 7
KEYWORDS 7
transmembrane protein
FEATURE 7
438-856 #domain glutamate receptor homology #label GRH\
507-524 #domain transmembrane #status predicted #label TM1\
565-584 #domain transmembrane #status predicted #label TM2\
608-626 #domain transmembrane #status predicted #label TM3\
637-663 #domain transmembrane #status predicted #label TM4\
821-841 #domain transmembrane #status predicted #label TM5\
415 #binding_site carbohydrate (Ash) (covalent) #status
predicted\
668 #binding_site phosphate (Ser) (covalent) #status
predicted
SUMMARY 7
#length 919 #molecular-weight 103991 #checksum 9539
Query Match 60.5%; Score 52; DB 2; Length 919;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 757 GGLIDSKGYG 766
      1:|||||:
      70 GALVDSKSYA 79
      1:|||||:
      7
RESULT 7
ENTRY 7
TITLE 7
ALTERNATE_NAMES 7
ORGANISM 7
DATE 7
ACCESSIONS 7
REFERENCE 7
AUTHORS 7
SUBMISSION 7
#cross-references EMBL:211716
#molecule_type mRNA
#residues 1-919 ##label SEE
#cross-references EMBL:211716
REFERENCE 7
AUTHORS 7
#journal 7
#title 7
#cross-references MUID:92153417
#accession 7
#molecule_type mRNA
#residues 32-117,'T',119-175,'T',177-209,'D',211,'DD',214-289,'P',
291-340,'P',342-354,'A',356-459,'DR',462-466,'ID',
469-919 ##label BET
##cross-references GB:M83552; NID:G204391; PID:G204392
##experimental_source cerebellum
COMMENT 7
Glutamate receptors form part of a major excitatory
neurotransmitter system of the brain and play roles in synapse
formation, in learning and memory, and in the pathogenesis of
neurodegenerative diseases.
CLASSIFICATION 7
KEYWORDS 7
transmembrane protein
FEATURE 7
438-856 #domain glutamate receptor homology #label GRH\
507-524 #domain transmembrane #status predicted #label TM1\
565-584 #domain transmembrane #status predicted #label TM2\
608-626 #domain transmembrane #status predicted #label TM3\
637-663 #domain transmembrane #status predicted #label TM4\
821-841 #domain transmembrane #status predicted #label TM5\
415 #binding_site carbohydrate (Ash) (covalent) #status
predicted\
668 #binding_site phosphate (Ser) (covalent) #status
predicted
SUMMARY 7
#length 919 #molecular-weight 103991 #checksum 9539
Query Match 60.5%; Score 52; DB 2; Length 919;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 757 GGLIDSKGYG 766
      1:|||||:
      70 GALVDSKSYA 79
      1:|||||:
      7
RESULT 7
ENTRY 7
TITLE 7
ALTERNATE_NAMES 7
ORGANISM 7
DATE 7
ACCESSIONS 7
REFERENCE 7
AUTHORS 7
SUBMISSION 7
#cross-references EMBL:211716
#molecule_type mRNA
#residues 1-919 ##label SEE
#cross-references EMBL:211716
REFERENCE 7
AUTHORS 7
#journal 7
#title 7
#cross-references MUID:92153417
#accession 7
#molecule_type mRNA
#residues 32-117,'T',119-175,'T',177-209,'D',211,'DD',214-289,'P',
291-340,'P',342-354,'A',356-459,'DR',462-466,'ID',
469-919 ##label BET
##cross-references GB:M83552; NID:G204391; PID:G204392
##experimental_source cerebellum
COMMENT 7
Glutamate receptors form part of a major excitatory
neurotransmitter system of the brain and play roles in synapse
formation, in learning and memory, and in the pathogenesis of
neurodegenerative diseases.
CLASSIFICATION 7
KEYWORDS 7
transmembrane protein
FEATURE 7
438-856 #domain glutamate receptor homology #label GRH\
507-524 #domain transmembrane #status predicted #label TM1\
565-584 #domain transmembrane #status predicted #label TM2\
608-626 #domain transmembrane #status predicted #label TM3\
637-663 #domain transmembrane #status predicted #label TM4\
821-841 #domain transmembrane #status predicted #label TM5\
415 #binding_site carbohydrate (Ash) (covalent) #status
predicted\
668 #binding_site phosphate (Ser) (covalent) #status
predicted
SUMMARY 7
#length 919 #molecular-weight 103991 #checksum 9539
Query Match 60.5%; Score 52; DB 2; Length 919;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 
```

```

Db 771 GGLIDSKGYG 780
   1-1-1111-1-1
QY 70 GALVDSKSYA 79

RESULT 9
ENTRY crtJ protein - Rhodobacter capsulatus
TITLE #formal_name Rhodobacter capsulatus
ORGANISM 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
DATE 30-Jun-1992

ACCESSIONS S17813
REFERENCE S17803
AUTHORS Burke, D.H.; Alberti, M.; Armstrong, G.A.; Hearst, J.E.
SUBMISSION submitted to the EMBL Data Library, November 1991.
DESCRIPTION The complete nucleotide sequence of the 46 kb photosynthesis
#accession S17813
#molecule_type DNA
#residues 1-469 #label EMB
#cross-references EMBL:Z11165
GENETICS crtJ
#gene #length 469 #molecular-weight 51331 #checksum 8204
SUMMARY
Query Match 59.3%; Score 51; DB 2; Length 469;
Best Local Similarity 42.9%; Pred. No. 6.60e+00;
Matches 6; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Db 308 DAGSAALVQGRSPA 321
   1-1-1111111-1-1
QY 66 DPSTGALVDSKSYA 79

RESULT 10
ENTRY S04886
TITLE mutL protein - Vibrio cholerae
ORGANISM #formal_name Vibrio cholerae
DATE 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
09-Sep-1997
ACCESSIONS S04886
REFERENCE S04886
AUTHORS Bera, T.K.; Ghosh, S.K.; Das, J.
JOURNAL Nucleic Acids Res. (1989) 17:6241-6251
TITLE Cloning and characterization of mutL and mutS genes of Vibrio
#cross-references MUID:89366654
#accession S04886
#molecule_type DNA
#residues 1-563 #label BER
#cross-references EMBL:X15438; NID:g48368; PID:g48369
GENETICS mutL
#gene #length 563 #molecular-weight 64858 #checksum 6294
SUMMARY
Query Match 59.3%; Score 51; DB 2; Length 563;
Best Local Similarity 63.6%; Pred. No. 6.60e+00;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 293 DPTGVLVERK 303
   1-1-111-1-1
QY 66 DPSTGALVDSK 76

RESULT 11
ENTRY A28063
TITLE glutamin-(asparagin-)-ase (EC 3.5.1.38) - Acinetobacter
ALTERNATE_NAMES calcoaceticus
ORGANISM glutaminase-asparaginase
#formal_name Acinetobacter calcoaceticus
DATE 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change
28-Apr-1993

ACCESSIONS A28063
REFERENCE Tanaka, S.; Robinson, E.A.; Appella, E.; Miller, M.; Ammon,
H.L.; Roberts, J.; Weber, I.T.; Wlodawer, A.
J. Biol. Chem. (1988) 263:8583-8591
#journal Structures of amidohydrolases. Amino acid sequence of a
#title glutaminase-asparaginase from Acinetobacter
glutaminasificans and preliminary crystallographic data for
an asparaginase from Erwinia chrysanthemi.
#cross-references MUID:88243706
#accession A28063
#molecule_type protein
#residues 1-331 #label TAN
#note the source is designated as Acinetobacter
glutaminasificans
CLASSIFICATION #superfamily asparaginase
KEYWORDS hydrolase
SUMMARY #length 331 #molecular-weight 35485 #checksum 4526
Query Match 58.1%; Score 50; DB 2; Length 331;
Best Local Similarity 55.6%; Pred. No. 1.06e+01;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 182 GTLVEGKPY 190
   1-1-111-1-1
QY 70 GALVDSKSY 78

RESULT 12
ENTRY JC5507
TITLE monocarboxylate transporter 3 - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change
18-Sep-1998
ACCESSIONS JC5507
REFERENCE JC5507
AUTHORS Yoon, H.; Fanelli, A.; Grollman, E.F.; Philp, N.J.
JOURNAL Biochem. Biophys. Res. Commun. (1997) 234:90-94
#title Identification of a unique monocarboxylate transporter (MCT3)
in retinal pigment epithelium.
#cross-references MUID:97312526
#accession JC5507
#status nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-542 #label YOO
#cross-references GB:U15685
#experimental_source retinal pigment epithelial cell
COMMENT This protein regulates lactate levels in the interphotoreceptor
space.
CLASSIFICATION #superfamily monocarboxylate transporter MCT2
KEYWORDS phosphoprotein
FEATURE 141
#binding_site phosphate (Thr) (covalent) (by casein
kinase II) #status predicted\
#binding_site phosphate (Thr) (covalent) (by CAMP- and
cGMP-dependent kinases) #status predicted\
#binding_site phosphate (Ser) (covalent) (by casein
kinase II) #status predicted\
#binding_site phosphate (Ser) (covalent) (by protein
kinase C) #status predicted\
SUMMARY #length 542 #molecular-weight 58085 #checksum 7501
Query Match 58.1%; Score 50; DB 2; Length 542;
Best Local Similarity 61.5%; Pred. No. 1.06e+01;
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 410 PSAGRLVDALKNY 422
   1-1-111-1-1
QY 67 PSTGALVDS-KSY 78

RESULT 13
ENTRY TVMVMD
#type complete

```

```

TITLE      protein-tyrosine kinase (EC 2.7.1.112) fms precursor - feline
            sarcoma virus (strain McDonough)
ORGANISM   #formal_name feline sarcoma virus
#note      host Felis sp. (cat)
DATE       27-Nov-1985 #sequence_revision 31-Dec-1991 #text_change
            13-Jun-1997
ACCESSIONS A00654
REFERENCE   Hampe, A.; Gobet, M.; Sherr, C.J.; Galibert, F.
#authors    Proc. Natl. Acad. Sci. U.S.A. (1984) 81:85-89.
#journal     Nucleotide sequence of the feline retroviral oncogene v-fms
#title       shows unexpected homology with oncogenes encoding
            tyrosine-specific protein kinases.
#cross-references MIM:84119469
#accession   A00654
#molecule_type DNA
#residues    1-941 #label HAM
COMMENT     This protein is synthesized as a gag-fms polyprotein.
GENETICS
#gene
CLASSIFICATION
#keywords    fms
            immunoglobulin homology; protein kinase homology
            ATP: autophosphorylation; glycoprotein; kinase-related
            transforming protein; magnesium; oncogene; phosphoprotein;
            phosphotransferase; receptor; transmembrane protein;
            tyrosine-specific protein kinase
FEATURE
1-23      #domain signal sequence #status predicted #label SIG\
24-941    #product protein-tyrosine kinase fms #status predicted
            #label HAM\
24-509    #domain extracellular #status predicted #label EXT\
35-86     #domain immunoglobulin homology #label IMM1\
120-179   #domain immunoglobulin homology #label IMM2\
217-280   #domain immunoglobulin homology #label IMM3\
316-381   #domain immunoglobulin homology #label IMM4\
410-484   #domain immunoglobulin homology #label IMM5\
510-534   #domain transmembrane #status predicted #label TM\
535-941   #domain intracellular #status predicted #label INT\
577-915   #domain protein kinase homology #label KIN\
585-593   #region protein kinase ATP-binding motif\
42-84,127-177,
224-278,417-482,
45,73,94,153,275,
286,302,335,410,
477,490   #disulfide_bonds #status predicted\
613,630,776
781,794   #binding_site carbohydrate (Asn) (covalent) #status
            predicted\
            #active_site Lys, Glu, Asp #status predicted\
            #binding_site magnesium (Asn, Asp) #status predicted\
SUMMARY    #length 941 #molecular_weight 104711 #checksum 3513
Query Match 58.1% Score 50; DB 1; Length 941;
Best Local Similarity 53.8%; Pred. No. 1.06e+01;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Db 852 NPYPGILVNSKFY 864
: | | | | | |
QY 66 DPSTGALVDSKSY 78
RESULT 14
ENTRY   TVHUMD #type complete
TITLE   macrophage colony-stimulating factor 1 receptor precursor -
            human
CONTAINS protein-tyrosine kinase (EC 2.7.1.112) csflr/fms
ORGANISM #formal_name Homo sapiens #common_name man
DATE     28-Dec-1987 #sequence_revision 31-Dec-1993 #text_change
            24-Oct-1997
ACCESSIONS S08123; A24533; I56672; I57648; I59083; I52772
REFERENCE   Hampe, A.; Shamooin, B.M.; Gobet, M.; Sherr, C.J.; Galibert,
#authors    F.
#journal     Oncogene Res. (1989) 4:9-17
#title       Nucleotide sequence and structural organization of the human

```

---

```

#cross-references MIM:89239490
#accession   S08123
#status      nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues    1-972 #label HAM
#cross-references GB:U63963; EMBL:X14720; NID:g1915975; PID:g1915976
#note        this sequence was submitted to the EMBL Data Library,
            March 1989
REFERENCE   A24533
#authors     Coussens, L.; Van Beveren, C.; Smith, D.; Chen, E.; Mitchell,
            R.L.; Isacke, C.M.; Verma, I.M.; Ullrich, A.
#journal     Nature (1986) 320:277-280
#title       Structural alteration of viral homologue of receptor
            proto-oncogene fms at carboxyl terminus.
#cross-references MIM:86175013
#accession   A24533
#molecule_type mRNA
#residues    1-53, 'A', '55-972 #label COU
#cross-references GB:J03149
#note        the authors translated the codon GCA for residue 54 as
            Pro
REFERENCE   I56672
#authors     Wheeler, E.F.; Rousset, M.F.; Hampe, A.; Walker, M.H.; Fried,
            V.A.; Look, A.T.; Rettenmier, C.W.; Sherr, C.J.
#journal     J. Virol. (1986) 59:224-233
#title       The amino-terminal domain of the v-fms oncogene product
            includes a functional signal peptide that directs synthesis
            of a transforming glycoprotein in the absence of feline
            leukemia virus gag sequences.
#cross-references MIM:86281820
#accession   I56672
#status      preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues    1-16 #label RES
#cross-references GB:M14002; NID:g182676; PID:g553292
REFERENCE   I57648
#authors     Vlasader, J.; Verma, I.M.
#journal     Mol. Cell. Biol. (1989) 9:1336-1341
#title       Differential transcription of exon 1 of the human c-fms gene
            in placental trophoblasts and monocytes.
#cross-references MIM:89261741
#accession   I57648
#status      preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues    1-16 #label RE2
#cross-references GB:M25786; NID:g349454; PID:g553224
REFERENCE   I59083
#authors     Browning, P.J.; Bunn, H.F.; Cline, A.; Shuman, M.; Nienhuis,
            A.W.
#journal     Proc. Natl. Acad. Sci. U.S.A. (1986) 83:7800-7804
#title       Replacement of COOH-terminal truncation of v-fms with c-fms
            sequences markedly reduces transformation potential.
#cross-references MIM:87017034
#accession   I59083
#status      translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues    874-972 #label RE3
#cross-references GB:M14193; NID:g182521; PID:g182522
REFERENCE   I52772
#authors     Nienhuis, A.W.; Bunn, H.F.; Turner, P.H.; Gopal, T.V.; Nash,
            W.G.; O'Brien, S.
#journal     Cell (1985) 42:421-428
#title       Expression of the human c-fms proto-oncogene in hematopoietic
            cells and its deletion in the 5q- syndrome.
#cross-references MIM:85282599
#accession   I52772
#status      preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues    244-295 #label RE4
#cross-references GB:M11067; NID:g182674; PID:g442423
GENETICS
#gene        GDB:CSF1R; FMS

```

```

##cross-references GDB:120600; OMIM:164770
#map_position 5q33.2-5q33.3
#introns 17/1; 103/1; 198/1; 243/3; 297/1; 361/2; 400/1; 440/2; 504/1;
542/3; 585/1; 620/1; 657/1; 711/2; 741/1; 773/3; 814/3;
852/1; 885/2; 921/3
CLASSIFICATION #superfamily macrophage colony-stimulating factor 1 receptor;
immunoglobulin homology; protein kinase homology
KEYWORDS ATP; autophosphorylation; glycoprotein; kinase-related
transforming protein; magnesium; phosphoprotein;
phosphotransferase; proto-oncogene; receptor; transmembrane
protein; tyrosine-specific protein kinase
FEATURE
1-23 #domain signal sequence #status predicted #label SIG\
24-972 #product macrophage colony-stimulating factor 1 receptor
#status predicted #label MAT\
24-512 #domain extracellular #status predicted #label EXT\
35-86 #domain immunoglobulin homology #label IMM1\
120-179 #domain immunoglobulin homology #label IMM2\
217-280 #domain immunoglobulin homology #label IMM3\
316-383 #domain immunoglobulin homology #label IMM4\
412-487 #domain immunoglobulin homology #label IMM5\
513-537 #domain transmembrane #status predicted #label TM\
538-972 #domain intracellular #status predicted #label INT\
580-917 #domain protein kinase homology #label KIN\
588-596 #region protein kinase ATP-binding motif\
42-84,127-177, #disulfide_bonds #status predicted\
224-278,419-485
45,73,153,240,275,
302,335,353,412,
428,480
#binding_site carbohydrate (Asn) (covalent) #status
predicted\
616,633,778 #active_site Lys, Glu, Asp #status predicted\
783,796 #binding_site magnesium (Asn, Asp) #status predicted
#length 972 #molecular-weight 107983 #checksum 2888
SUMMARY
Query Match 58.1%; Score 50; DB 1; Length 972;
Best Local Similarity 53.8%; Pred. No. 1.06e+01;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Db 854 NPYPGILVNSKFY 866
:| | | | |
Qy 66 DPSTGALVDSKSY 78

RESULT 15
ENTRY TVCTMD #type complete
TITLE macrophage colony-stimulating factor 1 receptor precursor -
cat
CONTAINS protein-tyrosine kinase (EC 2.7.1.112) csfir/fms
ORGANISM #formal name Felis silvestris catus #common_name domestic cat
DATE 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_Change
13-Jun-1997
ACCESSIONS A31636
REFERENCE Woolford, J.; McAuliffe, A.; Rohrschneider, L.R.
#authors Cell (1988) 55:965-977
#journal Activation of the feline c-fms proto-oncogene: multiple
#title alterations are required to generate a fully transformed
phenotype.
#cross-references MUID:89077553
#accession A31636
#molecule_type mRNA
#residues 1-980 #label WOO
##cross-references EMBL:X03663
GENETICS
#gene fms
CLASSIFICATION #superfamily macrophage colony-stimulating factor 1 receptor;
immunoglobulin homology; protein kinase homology
KEYWORDS ATP; autophosphorylation; glycoprotein; kinase-related
transforming protein; magnesium; phosphoprotein;
phosphotransferase; proto-oncogene; receptor; transmembrane
protein; tyrosine-specific protein kinase
FEATURE

```

```

1-23 #domain signal sequence #status predicted #label SIG\
24-980 #product macrophage colony-stimulating factor 1 receptor
#status predicted #label MAT\
24-509 #domain extracellular #status predicted #label EXT\
35-86 #domain immunoglobulin homology #label IMM1\
120-179 #domain immunoglobulin homology #label IMM2\
217-280 #domain immunoglobulin homology #label IMM3\
316-381 #domain immunoglobulin homology #label IMM4\
410-484 #domain immunoglobulin homology #label IMM5\
510-534 #domain transmembrane #status predicted #label TM\
535-980 #domain intracellular #status predicted #label INT\
577-915 #domain protein kinase homology #label KIN\
585-593 #region protein kinase ATP-binding motif\
42-84,127-177, #disulfide_bonds #status predicted\
224-278,417-482
45,73,94,153,275,
302,335,410,477,
490
#binding_site carbohydrate (Asn) (covalent) #status
predicted\
613,630,776 #active_site Lys, Glu, Asp #status predicted\
781,794 #binding_site magnesium (Asn, Asp) #status predicted
#length 980 #molecular-weight 108506 #checksum 7338
SUMMARY
Query Match 58.1%; Score 50; DB 1; Length 980;
Best Local Similarity 53.8%; Pred. No. 1.06e+01;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Db 852 NPYPGILVNSKFY 864
:| | | | |
Qy 66 DPSTGALVDSKSY 78

Search completed: Wed Sep 1 16:30:13 1999
Job time : 11 secs.

```

\*\*\*\*\*  
 W O R L D  
 \*\*\*\*\*

Release 3.1A John F. Collins, Blocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Wed Sep 1 16:28:57 1999; MasPar time 2.70 seconds  
 Tabular output not generated. 146.747 Million cell updates/sec

Title: >PCT-US99-13024-2  
 Description: (66-79) from PCTUS9913024.pep (11 of 12)  
 Perfect Score: 86  
 Sequence: 1 DPSTGALVDSKSYA 14

Scoring table: PAM 150  
 Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: swiss-prot37  
 1:swissprot

Statistics: Mean 25.367; Variance 25.400; scale 0.999

pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	52	60.5	285	1	GLK2_XENLA GLUTAMATE RECEPTOR, IO	1.22e+00
2	52	60.5	899	1	GLK2_MOUSE GLUTAMATE RECEPTOR, IO	1.22e+00
3	52	60.5	908	1	GLK2_HUMAN GLUTAMATE RECEPTOR, IO	1.22e+00
4	52	60.5	908	1	GLK2_RAT GLUTAMATE RECEPTOR, IO	1.22e+00
5	52	60.5	918	1	GLK1_HUMAN GLUTAMATE RECEPTOR, IO	1.22e+00
6	52	60.5	919	1	GLK3_RAT GLUTAMATE RECEPTOR, IO	1.22e+00
7	52	60.5	919	1	GLK3_HUMAN GLUTAMATE RECEPTOR, IO	1.22e+00
8	52	60.5	949	1	GLK1_RAT GLUTAMATE RECEPTOR, IO	1.22e+00
9	51	59.3	469	1	CRTJ_RHOCA CRTJ PROTEIN.	2.10e+00
10	51	59.3	563	1	MUTL_VIBCH GLUTAMINASE-ASPARAGIN	3.57e+00
11	50	58.1	331	1	ASPLQ_ACIGL GLUTAMINASE-ASPARAGIN	3.57e+00
12	50	58.1	972	1	KFMS_HUMAN MACROPHAGE COLONY STIM	3.57e+00
13	50	58.1	978	1	KFMS_FSVMD FMS TYROSINE KINASE TR	3.57e+00
14	50	58.1	980	1	KFMS_FELCA MACROPHAGE COLONY STIM	3.57e+00
15	49	57.0	235	1	YIP2_YEAST HYPOPHETICAL 27.2 KD P	6.03e+00
16	49	57.0	577	1	VGLE_PVRRI GLYCOPROTEIN GI PRECUR	6.03e+00
17	49	57.0	627	1	HEML_OPSTA 5-AMINOLEVULINIC ACID	6.03e+00
18	49	57.0	635	1	HEML_CHICK SERINE-REPEAT ANTIGEN	6.03e+00
19	49	57.0	989	1	SERA_PLAIFG INSULIN RECEPTOR SUBST	6.03e+00
20	49	57.0	1321	1	IR52_MOUSE PROTEIN SCK (FRAGMENT)	1.01e+01
21	48	55.8	428	1	SKC_HUMAN EXTRACELLULAR SERINE P	1.01e+01
22	48	55.8	448	1	PRTE_BACNO CRYPTIC PLASMIID PROTEI	1.67e+01
23	47	54.7	205	1	CPPC_NEIGO	

24	47	54.7	225	1	GTK1_RAT GLUTATHIONE S-TRANSFER	1.67e+01
25	47	54.7	434	1	YADA_YERPS INVASIN PRECURSOR (OUT	1.67e+01
26	47	54.7	450	1	DHE4_LACBI NADP-SPECIFIC GLUTAMAT	1.67e+01
27	47	54.7	457	1	DHE4_AGABI NADP-SPECIFIC GLUTAMAT	1.67e+01
28	47	54.7	473	1	VL2_HPV16 MINOR CAPSID PROTEIN L	1.67e+01
29	47	54.7	502	1	SYFB_YEAST PHENYLALANYL-CITRA SYNT	1.67e+01
30	47	54.7	522	1	GDS1_YEAST GDS1 PROTEIN.	1.67e+01
31	47	54.7	547	1	CITA_KLEPN SENSOR KINASE CITA (EC	1.67e+01
32	47	54.7	550	1	HEMA_IAME6 HEMAGGLUTININ PRECURSO	1.67e+01
33	47	54.7	583	1	T2F1_FLAOK TYPE IIS RESTRICTION E	1.67e+01
34	47	54.7	640	1	HEML_HUMAN 5-AMINOLEVULINIC ACID	1.67e+01
35	47	54.7	778	1	ACON_SCHPO ACONITATE HYDRATASE, M	1.67e+01
36	47	54.7	883	1	APCE_CYAPA PHYCOBILISOME LINKER P	1.67e+01
37	47	54.7	976	1	KFMS_MOUSE MACROPHAGE COLONY STIM	1.67e+01
38	47	54.7	978	1	KFMS_RAT MACROPHAGE COLONY STIM	1.67e+01
39	47	54.7	1356	1	KAB7_YEAST PROBABLE SERINE/THREON	1.67e+01
40	47	54.7	2233	1	RRPL_P13H4 RNA POLYMERASE BETA SU	1.67e+01
41	46	53.5	82	1	NUMM_MOUSE NADH-UBIQUINONE OXIDOR	2.75e+01
42	46	53.5	271	1	YKUB_YEAST HYPOPHETICAL 31.2 KD P	2.75e+01
43	46	53.5	343	1	FIBP_ADE07 FIBER PROTEIN.	2.75e+01
44	46	53.5	469	1	GLNA_STRCO GLUTAMINE SYNTHETASE (	2.75e+01
45	46	53.5	979	1	GLK5_MOUSE GLUTAMATE RECEPTOR, IO	2.75e+01

## ALIGNMENTS

RESULT	1					
ID	GLK2_XENLA	STANDARD;	PRT;	285	AA.	
AC	Q91755;					
DT	15-JUL-1998	(REL. 36, CREATED)				
DT	15-JUL-1998	(REL. 36, LAST SEQUENCE UPDATE)				
DE	GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 2 (GLUTAMATE RECEPTOR 6) (GLUR-					
DE	6) (FRAGMENT).					
GN	GRIK2 OR GLUR6.					
OS	XENOPUS LAEVIS (AFRICAN CLAWED FROG).					
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;					
CC	MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE-BRAIN:					
RX	MEDLINE; 96303825.					
RA	ISHIMARU H., KAMBOJ R., AMBROSINI A., HENLEY J.M., SOLOVIEV M.M.,					
RA	SUDAN H., ROSSIER J., ABUTIDZE K., RAMPERSAD V., USHERWOOD P.N.R.,					
RA	BATESON A.N., BARNARD E.A.;					
RT	"A unitary non-NMDA receptor short subunit from Xenopus: DNA cloning					
RT	and expression."					
RL	RECEPT. CHANNELS 4:31-49(1996).					
CC	FUNCTION: L-GLUTAMATE ACTS AS AN EXCITATORY NEUROTRANSMITTER AT					
CC	MANY SYNAPSES IN THE CENTRAL NERVOUS SYSTEM. THE POSTSYNAPTIC					
CC	ACTIONS OF GLU ARE MEDIATED BY A VARIETY OF RECEPTORS THAT ARE					
CC	NAMED ACCORDING TO THEIR SELECTIVE AGONISTS (BY SIMILARITY).					
CC	- - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.					
CC	- - SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.					
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-					
CC	tion between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch).					
CC	EMBL: X94116; E214777;					
DR	PFAM: PF00060; lig_chan; 1.					
DR	RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN;					
KW	TRANSMEMBRANE.					
FT	NON_TER 1 1					
FT	TRANSMEM 16 36					POTENTIAL.
FT	TRANSMEM 197 217					POTENTIAL.
FT	CARBOHYD 128 128					POTENTIAL.
SEQ	SEQUENCE 285 AA; 32041 MW; 5C7529B1 CRC32;					

Query Match 60.5%; Score 52; DB 1; Length 285;  
Best Local Similarity 60.0%; Pred. No. 1.22e+00;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 133 GGLIDSKGYG 142  
QY 70 GALVDSKSYA 79

RESULT 2  
ID GLK2\_MOUSE STANDARD; PRT; 889 AA.  
AC P39087; Q60933;  
DT 01-FEB-1995 (REL. 31, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 2 PRECURSOR (GLUTAMATE RECEPTOR  
DE 6) (GLUR-6) (GLUTAMATE RECEPTOR BETA-2) (GLUR BETA-2) (FRAGMENT).  
GN GRK2 OR GLUR6.  
OS MUS MUSCULUS (MUSCLE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
RP [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C; TISSUE-BRAIN;  
RX MEDLINE: 94083547.  
RA GREGOR P., O'HARA B.F., YANG X., UHL G.R.;  
RT "Expression and novel subunit isoforms of glutamate receptor genes  
RT Glur5 and Glur6.";  
RL NEUROREPORT 4:1343-1346(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92356790.  
RA MORITA T., SAKIMURA K., KUSHIYA E., YAMAZAKI M., MEGURO H.,  
RA ARAKI K., ABE T., MORI K.J., MISHINA M.;  
RT "Cloning and functional expression of a cDNA encoding the mouse beta  
RT 2 subunit of the kainate-selective glutamate receptor channel.";  
RL BRAIN RES. MOL. BRAIN RES. 14:143-146(1992).  
RN [3]  
RP SEQUENCE OF 584-695 FROM N.A., AND RNA EDITING.  
RC STRAIN-BALB/C;  
RX MEDLINE: 96312506.  
RA HERB A., HIGUCHI M., SPRENGEL R., SEEBURG P.H.;  
RT "Q/R site editing in kainate receptor Glur5 and Glur6 pre-mRNAs  
RT requires distant intronic sequences.";  
RL PROC. NATL. ACAD. SCI. U.S.A. 93:1875-1880(1996).  
RN [4]  
RP SEQUENCE OF 560-585 FROM N.A., AND RNA EDITING.  
RC TISSUE-BRAIN;  
RX MEDLINE: 93213505.  
RA KOEHLER M., BURNASHEV N., SAKMANN B., SEEBURG P.H.;  
RT "Determinants of Ca2+ permeability in both TM1 and TM2 of high  
RT affinity kainate receptor channels: diversity by RNA editing.";  
RL NEURON 10:491-500(1993).  
CC -!- FUNCTION: L-GLUTAMATE ACTS AS AN EXCITATORY NEUROTRANSMITTER AT  
CC MANY SYNAPSES IN THE CENTRAL NERVOUS SYSTEM. THE POSTSYNAPTIC  
CC ACTIONS OF GLU ARE MEDIATED BY A VARIETY OF RECEPTORS THAT ARE  
CC NAMED ACCORDING TO THEIR SELECTIVE AGONISTS. MAY BE INVOLVED IN  
CC THE TRANSMISSION OF LIGHT INFORMATION FROM THE RETINA TO THE  
CC HYPOTHALAMUS.  
CC -!- THE PRESENCE OF GLUTAMINE RESIDUE (NON-EDITED) DETERMINES CHANNELS  
CC WITH LOW CALCIUM PERMEABILITY, WHEREAS AN ARGININE RESIDUE  
CC (EDITED) DETERMINES A HIGHER CALCIUM PERMEABILITY ESPECIALLY IF  
CC THE PRECEDING SITES ARE FULLY EDITED (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- SUBUNIT: ASSEMBLES INTO A KAINATE-GATED HOMOMERIC CHANNEL THAT DOES  
CC NOT BIND AMPA. GRK2 ASSOCIATED TO GRK5 FORMS FUNCTIONAL  
CC CHANNELS. THAT CAN BE GATED BY AMPA (BY SIMILARITY).  
CC -!- TISSUE SPECIFICITY: MOST ABUNDANT IN THE CEREBELLUM AND THE  
CC HYPOTHALAMUS.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED DURING BRAIN DEVELOPMENT.  
CC EXPRESSION DROPS IN THE ADULT.  
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS GLUR6 BETA2 AND GLUR6-2 ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.

CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X66117; G312494; -;  
CC EMBL: D10054; D1001409; -;  
CC EMBL: U31443; G951154; -;  
CC PIR: S35792; S35792.  
CC PIR: A43954; A43954.  
CC MGI: 95815; GRK2.  
CC PFM: PFM0060; lig\_chan; 1.  
CC RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;  
CC TRANSMEMBRANE; ALTERNATIVE SPLICING; RNA EDITING.  
FT SIGNAL 1 31  
FT CHAIN 32 >889  
FT DOMAIN 32 561  
FT TRANSMEM 562 582  
FT TRANSMEM 601 621  
FT TRANSMEM 636 656  
FT TRANSMEM 820 840  
FT CARBOHYD 67 67  
FT CARBOHYD 73 73  
FT CARBOHYD 275 275  
FT CARBOHYD 378 378  
FT CARBOHYD 412 412  
FT CARBOHYD 423 423  
FT CARBOHYD 430 430  
FT CARBOHYD 546 546  
FT VARSPLIC 855 859  
FT VARSPLIC 870 >889  
FT VARIANT 567 567  
FT VARIANT 571 571  
FT VARIANT 621 621  
FT CONFLICT 611 611  
FT CONFLICT 849 889  
FT NON\_TER 889 889  
SQ SEQUENCE 889 AA; 100273 MW; 32E5917A CRC32;  
Query Match 60.5%; Score 52; DB 1; Length 889;  
Best Local Similarity 60.0%; Pred. No. 1.22e+00;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 756 GGLIDSKGYG 765  
QY 70 GALVDSKSYA 79

RESULT 3  
ID GLK2\_HUMAN STANDARD; PRT; 908 AA.  
AC Q13002;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 2 PRECURSOR (GLUTAMATE RECEPTOR  
DE 6) (GLUR-6) (EXCITATORY AMINO ACID RECEPTOR 4) (EAA4).  
GN GRK2 OR GLUR6.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RP [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-FETAL BRAIN;  
RX MEDLINE: 95236039.  
RA HOO K.H., NUTT S.L., FLETCHER E.J., ELLIOTT C.E., KORCZAK B.,  
RA DEVERILL R.M., RAMPERSAD V., FANTASKE R.P., KAMBOJ R.K.;

RT "Functional expression and pharmacological characterization of the human EAA4 (GluR6) glutamate receptor: a kainate selective channel subunit.";  
 RL RECEPT. CHANNELS 2:327-337(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94307730.  
 RA PASCHEN W., BLACKSTONE C.D., HUGANIR R.L., ROSS C.A.;  
 RT "Human GluR6 kainate receptor (GRIK2): molecular cloning, expression, polymorphism, and chromosomal assignment.";  
 RL GENOMICS 20:435-440(1994).  
 RN [3]  
 RP RNA EDITING.  
 RX MEDLINE: 95016699.  
 RA PASCHEN W., HEDREEN J.C., ROSS C.A.;  
 RT "RNA editing of the glutamate receptor subunits GluR2 and GluR6 in human brain tissue.";  
 RL J. NEUROCHEM. 63:1596-1602(1994).  
 RN [4]  
 RP RNA EDITING.  
 RC TISSUE-BRAIN;  
 RX MEDLINE: 95210645.  
 RA NUTT S.L., KAMBOJ R.K.;  
 RT "RNA editing of human kainate receptor subunits.";  
 RL NEUROREPORT 5:2625-2629(1994).  
 CC -!- FUNCTION: L-GLUTAMATE ACTS AS AN EXCITATORY NEUROTRANSMITTER AT MANY SYNAPSES IN THE CENTRAL NERVOUS SYSTEM. THE POSTSYNAPTIC ACTIONS OF GLU ARE MEDIATED BY A VARIETY OF RECEPTORS THAT ARE NAMED ACCORDING TO THEIR SELECTIVE AGONISTS. MAY BE INVOLVED IN THE TRANSMISSION OF LIGHT INFORMATION FROM THE RETINA TO THE HYPOTHALAMUS. THIS RECEPTOR BINDS DOMOATE > KAINATE > QUISQUALATE > 6-CYANO-7-NITROQUINOLAXINE-2,3-DIONE > L-GLUTAMATE - 6,7-DINITROQUINOLAXINE-2,3-DIONE > DIHYDROKAINATE.  
 CC -!- WITH THE PRESENCE OF GLUTAMINE RESIDUE (NON-EDITED) DETERMINES CHANNELS (EDITED) DETERMINES A HIGHER CALCIUM PERMEABILITY ESPECIALLY IF THE PRECEDING SITES ARE FULLY EDITED. THIS RECEPTOR IS NEARLY COMPLETELY EDITED IN ALL GRAY MATTER STRUCTURES (90% OF THE RECEPTORS), WHEREAS MUCH LESS EDITING OCCURS IN THE WHITE MATTER (10% OF THE RECEPTORS).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -!- SUBUNIT: ASSEMBLE INTO A KAINATE-GATED HOMOMERIC CHANNEL THAT DOES NOT BIND AMPA. GRIK2 ASSOCIATED TO GRIK5 FORMS FUNCTIONAL CHANNELS, THAT CAN BE GATED BY AMPA (BY SIMILARITY).  
 CC -!- TISSUE SPECIFICITY: EXPRESSION IS HIGHER IN CEREBELLUM THAN IN CEREBRAL CORTEX.  
 CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U16126; G790532;  
 DR MIM: 138244;  
 DR PFAM: PF00060; lig\_chan; 1.  
 KW RECEPTOR: POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL; TRANSMEMBRANE; RNA EDITING.  
 FT SIGNAL 1 31 POTENTIAL.  
 FT CHAIN 32 908 GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 2.  
 FT DOMAIN 32 561 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 562 582 POTENTIAL.  
 FT TRANSMEM 601 621 POTENTIAL.  
 FT TRANSMEM 636 656 POTENTIAL.  
 FT TRANSMEM 820 840 POTENTIAL.  
 FT CARBOHYD 67 67 POTENTIAL.  
 FT CARBOHYD 73 73 POTENTIAL.  
 FT CARBOHYD 275 275 POTENTIAL.  
 FT CARBOHYD 378 378 POTENTIAL.  
 FT CARBOHYD 412 412 POTENTIAL.

FT CARBOHYD 423 423 POTENTIAL.  
 FT CARBOHYD 430 430 POTENTIAL.  
 FT CARBOHYD 546 546 POTENTIAL.  
 FT VARIANT 567 567 I -> V (IN RNA EDITED VERSION).  
 FT VARIANT 571 571 Y -> C (IN RNA EDITED VERSION).  
 FT VARIANT 621 621 Q -> R (IN RNA EDITED VERSION).  
 FT CONFLICT 789 789 G -> S (IN REF. 2).  
 SQ SEQUENCE 908 AA; 102583 MW; D0058718 CRC32;  
 Query Match 60.5%; Score 52; DB 1; Length 908;  
 Best Local Similarity 60.0%; Pred. No. 1.22e+00;  
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Db 756 GGLIDSKGYG 765  
 QY 70 GALVDSKSYA 79  
 :|:|:|:|:  
 RESULT 4  
 ID GLK2\_RAT STANDARD; PRT; 908 AA.  
 AC P42260;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 2 PRECURSOR (GLUTAMATE RECEPTOR 6) (GLUR-6).  
 DE GRIK2 OR GLUR6.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-BRAIN;  
 RX MEDLINE: 91287799.  
 RA EGEJBERG J., BETTLER B., HERMANS-BORGMEYER I., HEINEMANN S.F.;  
 RT "Cloning of a cDNA for a glutamate receptor subunit activated by kainate but not AMPA.";  
 RL NATURE 351:745-748(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE: 93354697.  
 RA LOMELI H., WISDEN W., KOEHLER M., KEINAEEN K., SOMMER B., SEEBURG P.H.;  
 RT "High-affinity kainate and domoate receptors in rat brain.";  
 RL FEBS LETT. 307:139-143(1992).  
 RN [3]  
 RP SEQUENCE OF 560-585 FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE: 93213505.  
 RA KOEHLER M., BURNASHEV N., SAKMANN B., SEEBURG P.H.;  
 RT "Determinants of Ca2+ permeability in both TM1 and TM2 of high affinity kainate receptor channels: diversity by RNA editing.";  
 RL NEURON 10:491-500(1993).  
 CC -!- FUNCTION: L-GLUTAMATE ACTS AS AN EXCITATORY NEUROTRANSMITTER AT MANY SYNAPSES IN THE CENTRAL NERVOUS SYSTEM. THE POSTSYNAPTIC ACTIONS OF GLU ARE MEDIATED BY A VARIETY OF RECEPTORS THAT ARE NAMED ACCORDING TO THEIR SELECTIVE AGONISTS. MAY BE INVOLVED IN THE TRANSMISSION OF LIGHT INFORMATION FROM THE RETINA TO THE HYPOTHALAMUS. THIS RECEPTOR BINDS KAINATE > QUISQUALATE > GLUTAMATE. IT DOES NOT BIND AMPA WITHOUT COEXPRESSION WITH GRIK5.  
 CC -!- THE PRESENCE OF GLUTAMINE RESIDUE (NON-EDITED) DETERMINES CHANNELS (EDITED) DETERMINES A HIGHER CALCIUM PERMEABILITY ESPECIALLY IF THE PRECEDING SITES ARE FULLY EDITED. THIS RECEPTOR IS NEARLY COMPLETELY EDITED IN ALL GRAY MATTER STRUCTURES (90% OF THE RECEPTORS).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -!- SUBUNIT: ASSEMBLE INTO A KAINATE-GATED HOMOMERIC CHANNEL THAT DOES NOT BIND AMPA. GRIK2 ASSOCIATED TO GRIK5 FORMS FUNCTIONAL CHANNELS, THAT CAN BE GATED BY AMPA.  
 CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IS FOUND IN THE OLFACTORY LOBE, PIRIFORM CORTEX, DENTATE GYRUS, HIPPOCAMPUS, GRANULAR CELL



```

CC LAYER OF THE CEREBELLUM, AND IN CAUDATE-PUTAMEN.
CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z11548; G56282; -
CC DR EMBL; Z11715; G56280; -
CC DR PFAM; PF00060; lig_chan.1.
CC DR RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;
CC KW TRANSMEMBRANE; RNA EDITING.
CC NW SIGNAL 1 31
CC FT CHAIN 32 908 POTENTIAL.
CC FT DOMAIN 32 561 GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 2.
CC FT TRANSMEM 562 582 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 601 621 POTENTIAL.
CC FT TRANSMEM 636 656 POTENTIAL.
CC FT TRANSMEM 820 840 POTENTIAL.
CC FT CARBOHYD 67 67 POTENTIAL.
CC FT CARBOHYD 73 73 POTENTIAL.
CC FT CARBOHYD 275 275 POTENTIAL.
CC FT CARBOHYD 378 378 POTENTIAL.
CC FT CARBOHYD 412 412 POTENTIAL.
CC FT CARBOHYD 423 423 POTENTIAL.
CC FT CARBOHYD 430 430 POTENTIAL.
CC FT CARBOHYD 546 546 POTENTIAL.
CC FT VARIANT 567 567 I I -> C (IN RNA EDITED VERSION).
CC FT VARIANT 571 571 Y Y -> C (IN RNA EDITED VERSION).
CC FT VARIANT 621 621 Q -> R (IN RNA EDITED VERSION).
CC SQ SEQUENCE 908 AA; 102470 MW; CA86C64C CRC32;

Query Match 60.5%; Score 52; DB 1; Length 908;
Best Local Similarity 60.0%; Pred. No. 1.22e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 756 GGLIDSKGYG 765
|1:1||1:1|
QY 70 GALVDSKSYA 79

RESULT 5
ID ID GLK1_HUMAN STANDARD; PRT; 918 AA.
AC P39086; Q13001.
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 1 PRECURSOR (GLUTAMATE RECEPTOR
DE 5) (GLUR-5) (EXCITATORY AMINO ACID RECEPTOR 3) (EAA3).
GN GRIK1 OR GLUR5.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
CC [1]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=RETINA;
RX MEDLINE; 94083547.
RA GREGOR P., O'HARA B.F., YANG X., UHL G.R.;
RT "Expression and novel subunit isoforms of glutamate receptor genes
RT GluR5 and GluR6.";
RL NEUROREPORT 4:1343-1346(1993).
CC [2]
RN SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=FETAL BRAIN;
RX MEDLINE; 96172461.
RA KORCZAK B., NUTT S.L., FLETCHER E.J., HOO K.H., ELLIOTT C.F.,
RA RAMPERSAD V., MCWHINNIE E.A., KAMBOJ R.K.;
RT "cDNA cloning and functional properties of human glutamate receptor
RT EAA3 (GluR5) in homomeric and heteromeric configuration.";

```

RECEPT. CHANNELS 3:41-49(1995).

[3]

RN RNA EDITING.

RP TISSUE-BRAIN;

RC MEDLINE; 95210645.

RX NUTT S.L., KAMBOJ R.K.;

RA "RNA editing of human kainate receptor subunits.";

RT NEUROREPORT 5:2625-2629(1994).

RL

CC -1- FUNCTION: L-GLUTAMATE ACTS AS AN EXCITATORY NEUROTRANSMITTER AT MANY SYNAPSES IN THE CENTRAL NERVOUS SYSTEM. THE POSTSYNAPTIC ACTIONS OF GLU ARE MEDIATED BY A VARIETY OF RECEPTORS THAT ARE NAMED ACCORDING TO THEIR SELECTIVE AGONISTS. MAY BE INVOLVED IN THE TRANSMISSION OF LIGHT INFORMATION FROM THE RETINA TO THE HYPOTHALAMUS. THIS RECEPTOR BINDS DOMAIOE > KAINATE > L-GLUTAMATE = QUISQUALATE > CNQX > DNOX > AMPA > DIHYDROKAINATE > NMDA.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC -1- SUBUNIT: THE UNEDITED VERSION (Q) ASSEMBLES INTO A FUNCTIONAL KAINATE-GATED HOMOMERIC CHANNEL, WHEREAS THE EDITED VERSION (R) IS UNABLE TO PRODUCE CHANNEL ACTIVITY WHEN EXPRESSED ALONE. BOTH EDITED AND UNEDITED VERSIONS CAN FORM FUNCTIONAL CHANNELS WITH GRIK4 AND GRIK5 (BY SIMILARITY).

CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN THE CEREBELLUM AND THE SUPRACHIASMATIC NUCLEI (SCN) OF THE HYPOTHALAMUS.

CC -1- ALTERNATIVE PRODUCTS: AT LEAST TWO ISOFORMS (GLUR5-1D OR 1 AND EAA3A OR 2) ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.

CC

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; L15058; G455448; --

DR EMBL; U01625; G790530; --

DR MTM; 138245; --

DR PFAM; PF00060; lig\_chan; 1.

KW RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL; TRANSMEMBRANE; PHOSPHORYLATION; RNA EDITING; ALTERNATIVE SPLICING.

FT SIGNAL 1 30 POTENTIAL.

FT CHAIN 31 918 GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 1.

FT DOMAIN 31 576 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 577 597 POTENTIAL.

FT TRANSMEM 616 636 POTENTIAL.

FT TRANSMEM 654 674 POTENTIAL.

FT TRANSMEM 835 855 POTENTIAL.

FT CARBOHYD 68 68 POTENTIAL.

FT CARBOHYD 74 74 POTENTIAL.

FT CARBOHYD 276 276 POTENTIAL.

FT CARBOHYD 379 379 POTENTIAL.

FT CARBOHYD 428 428 POTENTIAL.

FT CARBOHYD 439 439 POTENTIAL.

FT CARBOHYD 446 446 POTENTIAL.

FT CARBOHYD 561 561 POTENTIAL.

FT MOD\_RES 725 725 PHOSPHORYLATION (BY PKC) (POTENTIAL).

FT MOD\_RES 761 761 PHOSPHORYLATION (BY PKC) (POTENTIAL).

FT VARIANT 636 636 Q -> R (IN RNA EDITED VERSION).

FT VARSPLIC 402 416 MISSING (IN ISOFORM 2).

FT VARSPLIC 870 918 AFFFFYGLAQKQTHPTNSTSGTFLSTDLCGKLIREGRIR KQSSVHTV -> CLSFNAIMEELGILSKNKKIKKSRKKG KSSFTSILTCRORTKQETVA (IN ISOFORM 2).

FT CONFLICT 281 281 R -> G (IN REF. 2).

FT SEQUENCE 918 AA; 103980 MW; 4A5251FB CRC32;

Query Match 60.5%; Score 52; DB 1; Length 918;

Best Local Similarity 60.0%; Pred. No. 1.22e+00;

Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 771 GGLDSKGYG 780

QY 70 GALYDSKSYA 79

```

Query Match      60.5%; Score 52; DB 1; Length 918;
Best Local Similarity 60.0%; Pred. No. 1.22e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 771 GGLIDSKGYG 780
Qy 70 GALVDSKSYA 79
  | | | | | |

```





```

FT  VARSPLIC 402 416 MISSING (IN GLURS-2).
FT  VARSPLIC 870 898 MISSING (IN GLURS-2B/GLURS-1 AND
FT                                     GLURS-2).
FT  VARSPLIC 870 871 KG -> HY (IN GLURS-2A).
FT  VARSPLIC 872 949 MISSING (IN GLURS-2A).
FT  CONFLICT 282 282 K -> L (IN REF. 2).
FT  CONFLICT 354 355 CA -> WR (IN REF. 2).
FT  CONFLICT 477 477 A -> G (IN REF. 2).
SQ  SEQUENCE 949 AA; 107840 MW; 7441125F CRC32;

Query Match          60.5%; Score 52; DB 1; Length 949;
Best Local Similarity 60.0%; Pred. No. 1.22e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db  771 GGLIDSKCYG 780
Qy  70 GALVDSKSYA 79

RESULT 9
ID  CRTJ-RHOCA STANDARD; PRT; 469 AA.
AC  P26167;
DT  01-MAY-1992 (REL. 22, CREATED)
DT  01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DE  CRTJ PROTEIN.
GN  CRTJ.
OS  RHODOBACTER CAPSULATUS (RHODOPSEUDOMONAS CAPSULATA).
OC  BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHODOBACTER GROUP;
OC  RHODOBACTER.
RN  [1]
RP  SEQUENCE FROM N.A.
RA  BURKE D.H., ALBERTI M., ARMSTRONG G.A., HEARST J.E.;
RL  SUBMITTED (NOV-1991) TO EMBL/GENBANK/DDBJ DATA BANKS.
RC  -!- PATHWAY: CAROTENOID AND CHLOROPHYLL BIOSYNTHESIS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; Z11165; G995679; -
DR  PIR; S17813; S17813.
DR  PFAM; PF00158; sigma54; 1.
KW  PHOTOSYNTHESIS; CHLOROPHYLL BIOSYNTHESIS; CAROTENOID BIOSYNTHESIS.
SQ  SEQUENCE 469 AA; 51363 MW; 6C3D4D39 CRC32;

Query Match          59.3%; Score 51; DB 1; Length 469;
Best Local Similarity 42.9%; Pred. No. 2.10e+00;
Matches 6; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Db  308 DAGSAALVQGRSFA 321
Qy  66 DPSTGALVDSKSYA 79

RESULT 10
ID  MUTL-VIBCH STANDARD; PRT; 563 AA.
AC  P11793;
DT  01-OCT-1989 (REL. 12, CREATED)
DT  01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DE  01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE  PROTEIN MUTL.
GN  MUTL.
OS  VIBRIO CHOLERAE.
OC  BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; VIBRIONACEAE; VIBRIO.
RN  [1]
RP  SEQUENCE FROM N.A.
RA  STRAIN-569B.
RC  MEDLINE; 89366654.
RX  -----

```

```

RA  BERA T.K., GHOSH S.K., DAS J.;
RT  "Cloning and characterization of mutL and mutS genes of Vibrio
RL  cholerae: nucleotide sequence of the mutL gene.";
CC  NUCLEIC ACIDS RES. 17:6241-6251(1989).
CC  -!- FUNCTION: MUTL IS INVOLVED IN THE METHYL DIRECTED REPAIR OF
CC  MISMATCHES IN DNA.
CC  -!- SIMILARITY: NONE WITH MUTL FROM OTHER BACTERIA.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; X15438; G48369; -
DR  PIR; S04886; S04886.
RW  DNA REPAIR.
SQ  SEQUENCE 563 AA; 64858 MW; 9F37C887 CRC32;

Query Match          59.3%; Score 51; DB 1; Length 563;
Best Local Similarity 53.6%; Pred. No. 2.10e+00;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db  293 DPRTGVLVERK 303
Qy  66 DPSTGALVDSK 76

RESULT 11
ID  ASPO-ACIGL STANDARD; PRT; 331 AA.
AC  P10172;
DT  01-MAR-1989 (REL. 10, CREATED)
DT  01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT  01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE  GLUTAMINASE-ASPARAGINASE (EC 3.5.1.38).
OS  ACINETOBACTER GLUTAMINASIFICANS.
OC  BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC  MORAXELLACEAE; ACINETOBACTER.
RN  [1]
RP  SEQUENCE.
RX  MEDLINE; 88243706.
RA  TANAKA S., ROBINSON E.A., APPELLA E., MILLER M., AMMON H.L.,
RA  ROBERTS J., WEBER I.T., WLODAWER A.;
RT  "Structures of amidohydrolases. Amino acid sequence of a glutaminase-
RT  asparaginase from Acinetobacter glutaminasificans and preliminary
RT  crystallographic data for an asparaginase from Erwinia
RT  chrysanthemi.";
RL  J. BIOL. CHEM. 263:8583-8591(1988).
RN  [2]
RP  SEQUENCE OF 1-60.
RX  MEDLINE; 78080774.
RA  HOLCENBERG J.S., ERICSSON L., ROBERTS J.;
RT  "Amino acid sequence of the diazooxonoleucine binding site of
RT  Acinetobacter and Pseudomonas 7A glutaminase-asparaginase enzymes.";
RL  BIOCHEMISTRY 17:411-417(1978).
RN  [3]
RP  X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
RA  LUBKOWSKI J., WLODAWER A., HOUSSET D., WEBER I.T., AMMON H.L.,
RA  MURPHY K.C., SWAIN A.L.;
RT  "Refined crystal structure of Acinetobacter glutaminasificans
RT  glutaminase-asparaginase.";
RL  ACTA CRYSTALLOGR. D 50:826-832(1994).
CC  -!- CATALYTIC ACTIVITY: L-GLUTAMINE + H(2)O = L-GLUTAMATE + NH(3).
CC  -!- SUBUNIT: HOMOTETRAMER.
CC  -!- ALSO CATALYZES: L-ASPARAGINE + H(2)O = L-ASPARTATE + NH(3).
DR  PIR; A28063; A28063.
DR  POB; 1AGX; 20-DEC-94.
DR  PROSITE; PS00144; ASN_GLN_ASE_1; 1.
DR  PROSITE; PS00917; ASN_GLN_ASE_2; 1.
DR  PFAM; PF00710; Asparaginase; 1.
KW  HYDROLASE; 3D-STRUCTURE.

```



```

QY      66 DPSTGALVDSKSY 78
RESULT 13
ID      KPMS.FSYMND      STANDARD;      PRT;      978 AA.
AC      P00345; Q86597;
DT      21-JUL-1986 (REL. 01, CREATED)
DT      01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT      15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE      FMS TYROSINE KINASE TRANSFORMING PROTEIN (EC 2.7.1.112).
GN      V-FMS.
OS      FELINE SARCOMA VIRUS (STRAIN MCDONOUGH).
OC      VIRUSES; RETROD VIRUSES; RETROVIRIDAE; MAMMALIAN TYPE C RETROVIRUSES.
RN      [1]
RX      MEDLINE; 84119469.
RA      HAMPE A., GOBET M., SHERR C.J., GALIBERT F.;
RT      "Nucleotide sequence of the feline retroviral oncogene v-fms shows
RT      unexpected homology with oncogenes encoding tyrosine-specific protein
RT      kinases.;"
RL      PROC. NATL. ACAD. SCI. U.S.A. 81:85-89(1984).
RN      [2]
RP      REVISIONS, SEQUENCE FROM N.A.
RX      MEDLINE; 92015516.
RA      SMOLA U., HENNIG D., HADWIGER-FANGMEIER A., SCHUETZ B., PFAFF E.,
RA      NIEMANN H., TAMURA T.;
RT      "Reassessment of the v-fms sequence: threonine phosphorylation of the
RT      COOH-terminal domain.;"
RL      J. VIROL. 65:6181-6187(1991).
CC      -!- FUNCTION: V-FMS IS DERIVED FROM THE RECEPTOR FOR COLONY
CC      STIMULATING FACTOR 1 (CSF-1).
CC      -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC      -!- THIS PROTEIN IS SYNTHESIZED AS A GAG-FMS POLYPEPTIDE.
CC      -!- SIMILARITY: BELONGS TO THE CSF-1/PDGF RECEPTOR FAMILY OF TYROSINE-
CC      PROTEIN KINASES.
CC      -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. THE
CC      EXTRACELLULAR DOMAIN CONTAINS FIVE IG-FOLD DOMAINS.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL; K01643; G323891; -
CC      EMBL; S59588; G237024; -
CC      PIR; A00654; TVMND.
CC      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC      PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC      PROSITE; PS00240; RECEPTOR_TYR_KIN_III; 1.
CC      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC      PFAM; PF00047; ig; 3.
CC      PFAM; PF00069; pkinase; 2.
CC      HSP; P11362; IFGL.
CC      POLYPEPTIDE; TRANSFORMING PROTEIN; TYROSINE-PROTEIN KINASE; ONCOGENE;
CC      TRANSFERASE; RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION;
CC      ATP-BINDING; IMMUNOGLOBULIN FOLD.
CC      DOMAIN 1 543 EXTRACELLULAR (POTENTIAL).
CC      TRANSMEM 544 588
CC      DOMAIN 569 978 CYTOPLASMIC (POTENTIAL).
CC      DOMAIN 58 138
CC      DOMAIN 141 231 IG-LIKE DOMAIN 1.
CC      DOMAIN 238 332 IG-LIKE DOMAIN 2.
CC      DOMAIN 333 431 IG-LIKE DOMAIN 3.
CC      DOMAIN 432 536 IG-LIKE DOMAIN 4.
CC      DOMAIN 432 536 IG-LIKE DOMAIN 5.
CC      DISULFID 76 118 POTENTIAL.
CC      DISULFID 161 211 POTENTIAL.
CC      DISULFID 258 312 POTENTIAL.
CC      DISULFID 451 516 POTENTIAL.
CC      DOMAIN 613 942 PROTEIN KINASE.
CC      NP_BIND 619 627 ATP (BY SIMILARITY).

```

---

```

FT      BINDING      647
FT      MOD_RES      841 841
FT      MOD_RES      973 973
FT      CARBOHYD      79 79
FT      CARBOHYD      107 107
FT      CARBOHYD      128 128
FT      CARBOHYD      187 187
FT      CARBOHYD      309 309
FT      CARBOHYD      320 320
FT      CARBOHYD      336 336
FT      CARBOHYD      369 369
FT      CARBOHYD      444 444
FT      CARBOHYD      524 524
FT      CARBOHYD      511 511
FT      CONFLICT      714 714
FT      CONFLICT      971 978
SQ      SEQUENCE 978 AA; 108491 MW; 2F1BCEFB CRC32;
Query Match      58.1%; Score 50; DB 1; Length 978;
Best Local Similarity 53.8%; Pred. No. 3.57e+00;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
DB      886 NPYPGILVNSKFY 898
QY      66 DPSTGALVDSKSY 78
RESULT 14
ID      KPMS.FELCA      STANDARD;      PRT;      980 AA.
AC      P13369;
DT      01-JAN-1990 (REL. 13, CREATED)
DT      01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT      01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE      MACROPHAGE COLONY STIMULATING FACTOR I RECEPTOR PRECURSOR (CSF-1-R)
DE      (EC 2.7.1.112) (FMS PROTO-ONCOGENE) (C-FMS).
GN      CSF1R OR FMS.
OS      FELIS SILVESTRIS CATUS (CAT).
OC      EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC      CARNIVORA; FISSIPEDIA; FELIDAE; FELIS.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 89077553.
RA      WOOLFORD J., MCAULIFFE A., ROHRSCHEIDER L.R.;
RT      "Activation of the feline c-fms proto-oncogene: multiple alterations
RT      are required to generate a fully transformed phenotype.;"
RL      CELL 55:965-977(1988).
CC      -!- FUNCTION: THIS PROTEIN IS THE RECEPTOR FOR CSF-1, IT IS A PROTEIN
CC      TYROSINE-KINASE TRANSMEMBRANE RECEPTOR.
CC      -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC      -!- SIMILARITY: BELONGS TO THE CSF-1/PDGF RECEPTOR FAMILY OF TYROSINE-
CC      PROTEIN KINASES.
CC      -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. THE
CC      EXTRACELLULAR DOMAIN CONTAINS FIVE IG-FOLD DOMAINS.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL; J03149; G163855; -
CC      PIR; A31636; TVTMD.
CC      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC      PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC      PROSITE; PS00240; RECEPTOR_TYR_KIN_III; 1.
CC      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC      PFAM; PF00047; ig; 3.
CC      PFAM; PF00069; pkinase; 2.
CC      HSP; P11362; IFGL.
CC      TRANSFERASE; GLYCOPROTEIN; TYROSINE-PROTEIN KINASE; RECEPTOR; TRANSMEMBRANE;
CC      PROTO-ONCOGENE; TYROSINE-PROTEIN KINASE; PHOSPHORYLATION; ATP-BINDING; SIGNAL;
CC      TRANSFERASE; GLYCOPROTEIN; PHOSPHORYLATION; ATP-BINDING; SIGNAL;

```

```
KW IMMUNOGLOBULIN FOLD. 19
FT SIGNAL 1
FT CHAIN 20 980
FT
FT
FT DOMAIN 20 509
FT TRANSMEM 510 535
FT DOMAIN 536 980
FT DOMAIN 24 104
FT DOMAIN 107 197
FT DOMAIN 204 298
FT DOMAIN 299 397
FT DOMAIN 398 502
FT DOMAIN 579 908
FT NP_BIND 585 593
FT BINDING 613 613
FT ACT_SITE 776 776
FT DISULFID 42 84
FT DISULFID 127 177
FT DISULFID 224 278
FT DISULFID 417 482
FT MOD_RES 807 807
FT CARBOHYD 45 45
FT CARBOHYD 73 73
FT CARBOHYD 94 94
FT CARBOHYD 153 153
FT CARBOHYD 275 275
FT CARBOHYD 286 286
FT CARBOHYD 302 302
FT CARBOHYD 335 335
FT CARBOHYD 410 410
FT CARBOHYD 477 477
FT CARBOHYD 490 490
SQ SEQUENCE 980 AA; 108506 MW; EAICFFB6 CRC32;

Query Match 58.1%; Score 50; DB 1; Length 980;
Best Local Similarity 53.8%; Pred. No. 3.57e+00;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 852 NPYPGILVNSKFY 864
QY 66 DPSTGALVDSKSY 78

RESULT 15
ID YIP2_YEAST STANDARD; PRT; 235 AA.
AC P40455;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 27.2 KD PROTEIN IN IMP2-DNA43 INTERGENIC REGION.
GN YIL152W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
RA CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,
RA GENTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
RA LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
RA RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
RA WALSH S.V., WHITEHEAD S.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL; 247047; G763194; -.
DR EMBL; 238059; G557771; -.
DR PIR; S48382; S48382.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 235 AA; 27201 MW; E0E5455D CRC32;

Query Match 57.0%; Score 49; DB 1; Length 235;
Best Local Similarity 46.2%; Pred. No. 6.03e+00;
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 207 DPSASLVDSRSF 219
QY 66 DPSTGALVDSKSY 78

Search completed: Wed Sep 1 16:29:07 1999
Job time : 10 secs.
```

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.

```

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:17:15 1999; MasPar time 6.29 Seconds
        314.294 Million cell updates/sec
Tabular output not generated.

```

```

Title:
Description:
Perfect Score:
Sequence:
Scoring table:

```

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

```
Database:
a-genes35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39
```

Statistics: Mean 27.254; Variance 116.070; scale 0.235

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARY

Result No.	Score	Match	Query %	Length	DB	ID	Description	Pred. No.
1	300	47.3	65.19	R97371			Phage T4 ORF X gene pr	1.86e-18
2	96	15.1	1693.35	R76368			Hepatitis E virus hol	3.01e+00
3	92	14.5	520.1	P94617			Neutral protease enco	6.20e+00
4	92	14.5	521.3	P10009			Sequence of neutral p	6.20e+00
5	87	13.7	242.19	P95000			Bacillus subtilis lic	1.51e+01
6	87	13.7	623.35	P57573			Amino acid sequence o	1.51e+01
7	86	13.6	239.1	R06621			Hybrid (1,3-1,4)-pre-	1.80e+01
8	86	13.6	1693.34	R71209			Protein encoded by OR	1.80e+01
9	86	13.6	1693.19	R91813			Hepatitis E virus str	1.80e+01
10	86	13.6	1693.10	R51264			HEV strain protein en	1.80e+01
11	86	13.6	1693.37	W81519			Hepatitis E virus (HE	1.80e+01
12	86	13.6	1693.36	W80196			Protein encoded by OR	1.80e+01
13	85	13.4	633.35	P57574			Amino acid sequence o	2.14e+01
14	84	13.2	1693.3	R146618			Protein encoded by OR	2.55e+01
15	83	13.1	613.36	W73122			A. thaliana ethylene	3.04e+01
16	83	13.1	613.15	R74632			QTR ethylene respons	3.04e+01



```

AC W76368;
DT 03-DEC-1998 (first entry)
DE Hepatitis E virus hollow particle protein #1.
KW Hollow particle protein; virus; antibody; detection; immunoassay;
KW infection.
OS Hepatitis virus.
FH Key Location/Qualifiers
FT Protein 1..1693
FT /note="Partial sequence"
PN J10234383-A.
PD 08-SEP-1998.
PF 28-FEB-1997; 062445.
PR 28-FEB-1997; JP-062445.
PA (DENK-) DENKA SEIKEN KK.
PA (KOKU-) KOKURITSU YOBO EISEI KENKYUSHO.
DR WPI: 98-535037/46.
DR N-PSDB: V61887.
PT Hepatitis E virus hollow particle poly:peptide(s) and nucleic acids
PT encoding it - useful for more accurate detection of HEV in samples,
PT using immuno-assays and nucleic acid hybridisation
PS Claim 10; Page 17-24; 29pp; Japanese.
CC This sequence represents a Hepatitis E viral hollow particle protein.
CC Infection in samples, e.g. by immuno-assay based techniques, and the
CC nucleic acid can be used for the same in nucleic acid hybridisation
CC assays. The polypeptides and nucleic acids allow more accurate
CC detection of HEV than previously possible.
SQ Sequence 1693 AA;

Query Match 15.1%; Score 96; DB 35; Length 1693;
Best Local Similarity 26.9%; Pred. No. 3.01e+00;
Matches 18; Conservative 21; Mismatches 25; Indels 3; Gaps 3;

DB 572 frtsfdvgavleangperynlfsdsgstmaagpfsptyaasaglevryvaagldhrav 631
QY 8 FGQYVOTPFLESNSVRKISAGSPLSTAGPSYVKFQDNVGSQF-F-SAGLHLR-V 64

DB 632 fapgvsp 638
QY 65 FDPSTGA 71

RESULT 3
ID P94617 standard; protein: 520 AA.
AC P94617;
DT 21-JUN-1990 (first entry)
DE Neutral protease encoded by npr gene.
KW Protease; expression systems; subtilin; neutral protease; ds.
OS Bacillus amyloliquefaciens.
FH Key Location/Qualifiers
FT Protein 221..520
FT /note="Mature"
FT region 27..520
FT /note="PRO-"
PN US4801537-A.
PD 31-JAN-1989.
PF 29-MAR-1985; 717800.
PR 29-MAR-1985; US-717800.
PA (GENE-) Genex Corp.
PI Nagarajan V, Rhodes CS, Banner CDB;
DR WPI: 89-053639/07.
DR N-PSDB: N91114.
PT Vectors for expression of polypeptide(s) in Bacillus -
PT contg. promoter and regulatory regions which control expression
PT and secretion of protease(s) in Bacillus.
PS Disclosure; p; English.
CC Claimed replicon comprises a promoter and regulatory regions, capable of
CC expressing alkaline and neutral protease genes.
SQ Sequence 520 AA;

Query Match 14.5%; Score 92; DB 1; Length 520;
Best Local Similarity 27.8%; Pred. No. 6.20e+00;
Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps 2;

```

```

DB 239 lssesgkyvirdlskptgtqtlitydlnreynlp-gtlvsstnqftssgraa 291
QY 36 LSTAGPSYV-KFQDNPVGSQTFSGAGLHLRVFDPSTGALVDSKSYAFSTSDTTS 88

RESULT 4
ID P51009 standard; Protein; 521 AA.
AC P51009;
DT 01-DEC-1991 (first entry)
DE Sequence of neutral protease encoded by the npr[Bamp] gene.
KW Bacillus expression vector; secretion vector.
OS Bacillus amyloliquefaciens.
FH Key Location/Qualifiers
FT peptide 1..27
FT /label= signal
FT peptide 28..221
FT /label= Pro sequence
FT protein 222..521
FT misc_difference 1
FT /label= fMet
PN EP-133756-A.
PD 06-MAR-1985.
PF 06-JUL-1984; 304662.
PR 06-JUL-1983; US-511198.
PR 08-JUN-1984; US-618902.
PR 29-MAR-1985; US-717800.
PA (GENE-) GENEX CORP.
PI Nagarajan V, Banner CDB, Rhodes CS;
DR WPI: 85-057299/10.
DR N-PSDB: N50542.
PT Replicable plasmidic expression vector - for transformation of
PT Bacillus to direct expression of poly:peptide
PS Disclosure; Fig 4; 3pp; English.
CC The inventors claim a vector comprising a replicable plasmid
CC containing the promoter and regulatory region of a gene selected
CC from apr[Bamp] and npr[Bamp], for transformation of Bacillus to
CC direct expression of polypeptide.
SQ Sequence 521 AA;

Query Match 14.5%; Score 92; DB 3; Length 521;
Best Local Similarity 27.8%; Pred. No. 6.20e+00;
Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps 2;

DB 240 lssesgkyvirdlskptgtqtlitydlnreynlp-gtlvsstnqftssgraa 292
QY 36 LSTAGPSYV-KFQDNPVGSQTFSGAGLHLRVFDPSTGALVDSKSYAFSTSDTTS 88

RESULT 5
ID P95000 standard; Protein; 242 AA.
AC P95000;
DT 12-FEB-1997 (first entry)
DE Bacillus subtilis lichenase.
KW Beer production; fermentation; barley; beta-glucan; hydrolysis;
KW lichenase.
OS Bacillus subtilis Y-25.
PN J01067181-A.
PD 13-MAR-1989.
PF 08-SEP-1987; 224615.
PR 08-SEP-1987; JP-224615.
PA (ASAK ) ASAKI BREWERIES KK.
DR WPI: 89-119863/16.
DR N-PSDB: N95000.
PT Recombinant plasmid used in beer prodn. - obtd. by integrating
PT lichenase gene derived from Bacillus subtilis, into vector
PS Disclosure; Fig 4; 7pp; Japanese.
CC The lichenase gene from Bacillus subtilis Y-25 is used for
CC transforming Bacillus hosts so that they show increased lichenase
CC expression. The recombinant lichenase enzyme produced by the
CC transformants is useful in beer production for decomposing beta-
CC glucan from barley.
SQ Sequence 242 AA;

```

Query Match 13.7%; Score 87; DB 19; Length 242;  
Best Local Similarity 27.3%; Pred. No. 1.51e+01;  
Matches 15; Conservative 19; Mismatches 18; Indels 3; Gaps 3;

Db 77 laitspsynkfcdgcnsvqtygygyerm-kpakntgivssfftytptdtp 130  
QY 36 LSTAGPSYVKFQ-DNPVGSQTFSAGLH-LRVDFPSTGALVDSKSYAFSTSDNTS 88

RESULT 6  
ID W75773 standard; Protein; 623 AA.  
AC W75773;  
DT 02-DEC-1998 (first entry)  
DE Amino acid sequence of lepidoteran-active HD573 toxin.  
KW HD573 toxin; PCR; primer; amplification; Bacillus thuringiensis; probe;  
KW lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;  
KW Helicoverpa zea; hybridisation.  
OS Bacillus thuringiensis.  
PN WO9840490-A1.  
PD 17-SEP-1998.  
PF 13-MAR-1998; U05081.  
PR 13-MAR-1997; US-040512.  
PA (MYCO ) MYCOGEN CORP.  
PI Muller-Cohn J, Narva KE, Schnepf HE;  
DR WPI: 98-506734/43.  
DR N-PSDB: V52610.  
PT New insecticidal Bacillus thuringiensis toxins - useful for  
PT controlling lepidopteran pests, especially Ostrinia nubilalis,  
PT Heliothis virescens and Helicoverpa zea  
PS Claim 14; Pages 28-30; 50pp; English.  
CC This is the amino acid sequence of a novel Bacillus thuringiensis toxin  
CC used in the method of the invention, to control lepidopteran pests.  
CC The new toxins are useful as pesticides, especially for the control of  
CC Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The  
CC polynucleotide coding sequences are useful for recombinant expression  
CC of the toxins and the primers, together with probes derived from the  
CC new sequences, are useful for the identification and characterisation  
CC of novel genes that encode pesticidal toxins.  
SQ Sequence 623 AA;

Query Match 13.7%; Score 87; DB 35; Length 623;  
Best Local Similarity 18.2%; Pred. No. 1.51e+01;  
Matches 12; Conservative 26; Mismatches 26; Indels 2; Gaps 2;

Db 502 fisekyngqdsrlfslsttarytlrgnqsynlyrvssigsstirvtingryvtanv 561  
QY 12 YVQTFPLSESNVRYKIS-IAGSCLPSTAGPSYVKF-QDNPVGSQTFSAGLHLRVDFPST 69

Db 562 nttnnn 567  
QY 70 GALVDS 75

RESULT 7  
ID R06621 standard; protein; 239 AA.  
AC R06621;  
DT 09-JAN-1991 (first entry)  
DE Hybrid (1.3-1.4)-pre-beta-glucanase.  
KW Hybrid pre-beta-glucanase; glucans; beer; animal feed; poultry.  
OS Bacillus amyloliquefaciens, Bacillus macerans.  
FH Key Location/Qualifiers  
FT domain 1..129  
FT /label-amino terminal of beta-amyloliquefaciens  
FT domain 133..236  
FT /label-carboxyl-terminal of B.macerans  
PN WO9009436-A.  
PD 23-FEB-1990.  
PF 16-FEB-1990; DK0044.  
PR 16-FEB-1989; DD-325800.  
PR 04-AUG-1989; DK-003848.  
PA (CARL-) CARLSBERG A/S.  
PA (DEAK ) ARAD WISSENSCHAFT DDR.

PI Borriass R, Hofemeister J, Thomsen KK, Olsen O, Vonwettstein D;  
DR WPI: 90-275129/36.  
DR N-PSDB: Q05832.  
PT New thermostable (1.3-1.4)-beta-glucanase - prep'd. using hybrid  
PT gene obt'd. using Bacillus amyloliquefaciens and B.macerans genes  
PS Disclosure; page 26; 84pp; English.  
CC This hybrid protein is encoded by the beta-glucanase-H1 gene.  
CC Following processing of the signal peptide the mature protein  
CC is produced, comprising the amino terminus of the amylolique-  
CC faciens beta-glucanase and the carboxyl-terminal half of the  
CC B.macerans beta-glucanase. This hybrid protein is thermostable  
CC and hydrolyses beta-glycosidic linkages in (1,3-1,4)-beta-glucans.  
CC Reducing sugars are obt'd. at high temps. and thus this enzyme can  
CC be used in the mfr. of food prods., esp. beer and animal feed (eg  
CC for feeding poultry). See also Q05833.  
SQ Sequence 239 AA;

Query Match 13.6%; Score 86; DB 1; Length 239;  
Best Local Similarity 25.5%; Pred. No. 1.80e+01;  
Matches 14; Conservative 20; Mismatches 18; Indels 3; Gaps 3;

Db 74 laitspsynkfcdgcnsvqtygygyerm-kpakntgivssfftytptdtp 127  
QY 36 LSTAGPSYVKFQ-DNPVGSQTFSAGLH-LRVDFPSTGALVDSKSYAFSTSDNTS 88

RESULT 8  
ID W71209 standard; Protein; 1693 AA.  
AC W71209;  
DT 30-OCT-1998 (first entry)  
DE Protein encoded by ORF 1 of the Burmese isolate of ET-NANB.  
KW Enterically transmitted nonA/nonB hepatitis virus; identification;  
KW HEV; ET-NANB; detection; vaccine.  
OS Hepatitis virus.  
FH Key Location/Qualifiers  
FT Misc\_difference 154  
FT /note= "not specified"  
FT Misc\_difference 1514  
FT /note= "not specified"  
FT Misc\_difference 1552  
FT /note= "not specified"  
PN US5789559-A.  
PD 04-AUG-1998.  
PF 25-JUN-1994; 279823.  
PR 05-APR-1991; US-681078.  
PR 17-JUN-1988; US-208997.  
PR 11-APR-1989; US-336672.  
PR 19-JUN-1989; US-367486.  
PR 13-OCT-1989; US-420921.  
PR 05-APR-1990; US-505888.  
PR 25-JUL-1994; US-279823.  
PA (GENE-) GENELABS TECHNOLOGIES INC.  
PI Bradley DW, Fry KE, Krawczynski KZ, Reyes GR, Tam A,  
PI Yarbough PO;  
DR WPI: 98-446186/38.  
DR N-PSDB: V54729.  
PT Hepatitis E virus DNA - useful for e.g. virus detection and viral  
PT protein production  
PS Disclosure; Columns 55-64; 45pp; English.  
CC W71209-11 represent the proteins encoded by the open reading frames  
CC (ORFs) of the DNA sequence of the Burmese isolate of an enterically  
CC transmitted nonA/nonB viral hepatitis agent (ET-NANB). The nucleic  
CC acid sequence may be used for identifying and sequencing the entire  
CC viral agent (also referred to as HEV), detecting ET-NANB in  
CC infected samples, e.g. by specific amplification of virus-derived DNA  
CC sequences and for producing recombinant viral proteins for use in  
CC vaccines.  
SQ Sequence 1693 AA;

Query Match 13.6%; Score 86; DB 34; Length 1693;  
Best Local Similarity 25.4%; Pred. No. 1.80e+01;  
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

```

Db      572 frtsfdvgavlelngperhnlfsdaagstnaagpfsltyaasaaglevrvyvaagldhrav 631
QY      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
8 FGQGYVTPFLSESNVSRYKISAGSCPLSTAGPSYVKFQDNPVGQT-F-SAGLHLR-V 64

Db      632 fapgvsp 638
QY      ||::|
65 FDPSTGA 71

RESULT 9
ID R01813 standard; Protein; 1693 AA.
AC R01813;
DT 26-NOV-1996 (first entry)
DE Hepatitis E virus strain SAR-55 ORF-1.
KW Hepatitis E virus; HEV; SAR-55 strain; enteric transmission;
KW structural region; antigen; detection; antibody; vaccine;
KW immunisation; infection.
OS Hepatitis E virus.
FH Key Location/Qualifiers
FT misc_difference 1238
FT FT /note= "corresponding codon CAG"
FT FT /misc_difference 1455..1693
FT FT /note= "10 bp nucleic acid sequence TGCTNTTYGA
has to be inserted between nucleotides
4390..4391 of T27394 before these amino
acid residues can be decoded"
FT FT WO9610580-A2.
PN PD 11-OCT-1996.
PF PF 03-OCT-1995; UI3102.
PR PR (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA Emerson SU, Purcell RH, Tsarev SA;
PI WPI: 96-209320/21.
DR DR N-PSDB: T27394.
PT Isolated and purified hepatitis E virus strain SAR-55 DNA - encodes
PT antigenic protein useful in diagnosis, prophylaxis and treatment of
PT hepatitis E virus infection
PS Disclosure; Pages 9-13; 12lpp; English.
CC The present sequence is the protein prod. of ORF-1 from the
CC hepatitis E virus (HEV) strain SAR-55, which was implicated in an
CC enterically transmitted non-A, non-B hepatitis in Pakistan. The
CC protein encoded by the structural region of the virus (i.e. ORF-2),
CC which is capable of forming HEV like particles, is useful for the
CC detection of HEV antibodies (pref. IgG or IgM) in blood, plasma,
CC sera, cerebrospinal fluid, tissue, urine or pleural fluid. The
CC protein, and anti-HEV antibodies generated using the protein, can
CC also be used in vaccines for immunising an animal against HEV
CC infection. The protein is identified as a band of greater than
CC 50 kd following SDS-PAGE of cell lysates of insect cells infected
CC with a HEV ORF-2 contg. baculovirus, i.e. the claimed recombinant
CC expression vectors pPIC9-1779, -1780 and -1781.
SQ Sequence 1693 AA;

Query Match 13.6%; Score 86; DB 19; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.80e+01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps

Db      572 frtsfdvgavlelngperhnlfsdaagstnaagpfsltyaasaaglevrvyvaagldhrav 631
QY      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
8 FGQGYVTPFLSESNVSRYKISAGSCPLSTAGPSYVKFQDNPVGQT-F-SAGLHLR-V 64

Db      632 fapgvsp 638
QY      ||::|
65 FDPSTGA 71

RESULT 10
ID R51264 standard; Protein; 1693 AA.
AC R51264;
DT 21-OCT-1994 (first entry)
DE HEV strain protein encoded by ORF-1
KW Hepatitis E virus; HEV; strain SAR-55; open reading frame; ORF;
KW antibody; detection; diagnosis; primates; stool suspension.
```

CC HEV, and for monitoring the progression of such disease. Such methods are  
 CC also useful for monitoring the efficacy of therapeutic agents during the  
 CC course of treatment of HEV infection and disease in a mammal. The  
 CC antibodies can be used for detection or for passive immunisation of  
 CC mammals.  
 SQ Sequence 1693 AA;

Query Match 13.6%; Score 86; DB 37; Length 1693;  
 Best Local Similarity 25.4%; Pred. No. 1.80e+01;  
 Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;  
 Db 572 ftsfdvgavltngperhnlfsdagsqstmaagpfslytaasaaglevryvaagldhrav 631  
 QY 8 FGGYVOTPFLESNSVRKYISAGSCLSTAGPSYVKFQDNVPGSQT-F-SAGLHLR-V 64  
 Db 632 fapgvsp 638  
 QY 65 FDPSTGA 71

RESULT 12  
 ID W80196 standard; Protein: 1693 AA.  
 AC W80196;  
 DT 23-DEC-1998 (first entry)  
 DE Protein encoded by ORF1 of ET-NANB (HEV) Burma strain DNA sequence.  
 KW Enterically transmitted non A non B hepatitis virus; ET-NANB;  
 KW Hepatitis E virus; HEV; Burma HEV isolate; vaccine;  
 KW diagnostic probe.  
 OS Non A non B Hepatitis virus.  
 PN US5824649-A.  
 PD 20-OCT-1998.  
 PF 07-JUN-1995; 475807.  
 PR 25-JUL-1994; US-279823.  
 PR 17-JUN-1988; US-208997.  
 PR 11-APR-1989; US-336672.  
 PR 16-JUN-1989; US-367486.  
 PR 13-OCT-1989; US-420921.  
 PR 05-JUL-1990; US-505888.  
 PR 07-JUN-1995; US-475807.  
 PA (GENE-) GENELABS TECHNOLOGIES INC.  
 PI Bradley DW, Fry KE, Krawczynski KZ, Reyes GR, Tam A,  
 PI Yarbough PO;  
 DR N-PSDB: V66321.  
 DR N-PSDB: V66321.

PT Hepatitis E virus proteins - useful for diagnosis or vaccine  
 PT production the virus  
 PS Claim 22: Columns 57-66; 47pp; English.  
 CC W80196-98 are encoded by the genome of the Burma strain of  
 CC enterically transmitted non A non B hepatitis virus (ET-NANB)  
 CC (hepatitis E virus (HEV)). The specification describes an isolated  
 CC protein which is specifically immunoreactive with antibodies present  
 CC in individuals infected with HEV and encoded by a sequence contained  
 CC in an open reading frame (ORF) of an HEV genome. The genome has a  
 CC sequence that is more than 70% identical to the ORF1 sequence from  
 CC Burma HEV isolate. The protein is used as a vaccine and a diagnostic  
 CC probe for ET-NANB.  
 SQ Sequence 1693 AA;

Query Match 13.6%; Score 86; DB 36; Length 1693;  
 Best Local Similarity 25.4%; Pred. No. 1.80e+01;  
 Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 ftsfdvgavltngperhnlfsdagsqstmaagpfslytaasaaglevryvaagldhrav 631  
 QY 8 FGGYVOTPFLESNSVRKYISAGSCLSTAGPSYVKFQDNVPGSQT-F-SAGLHLR-V 64  
 Db 632 fapgvsp 638  
 QY 65 FDPSTGA 71

RESULT 13  
 ID W75774 standard; Protein: 633 AA.

AC W75774;  
 DT 02-DEC-1998 (first entry)  
 DE Amino acid sequence of lepidoteran-active HD525 toxin.  
 KW HD525 toxin; PCR; primer; amplification; Bacillus thuringiensis; probe;  
 KW lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;  
 KW Helicoverpa zea; hybridisation.  
 OS Bacillus thuringiensis.  
 PN WO9840490-A1.  
 PD 17-SEP-1998.  
 PF 13-MAR-1998; U05081.  
 PR 13-MAR-1997; US-040512.  
 PA (MYCO ) MYCOGEN CORP.  
 PI Muller-Cohn J, Narva KE, Schnepf HE;  
 PI WPI: 98-506734/43.  
 DR N-PSDB: V52611.  
 DT New insecticidal Bacillus thuringiensis toxins - useful for  
 PT controlling lepidopteran pests, especially Ostrinia nubilalis,  
 PT Heliothis virescens and Helicoverpa zea  
 PS Claim 14; Pages 32-34; 50pp; English.  
 CC This is the amino acid sequence of a novel Bacillus thuringiensis toxin  
 CC used in the method of the invention, to control lepidopteran pests.  
 CC The new toxins are useful as pesticides, especially for the control of  
 CC Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The  
 CC polynucleotide coding sequences are useful for recombinant expression  
 CC of the toxins and the primers, together with probes derived from the  
 CC new sequences, are useful for the identification and characterisation  
 CC of novel genes that encode pesticidal toxins.  
 SQ Sequence 633 AA;

Query Match 13.4%; Score 85; DB 35; Length 633;  
 Best Local Similarity 23.7%; Pred. No. 2.14e+01;  
 Matches 14; Conservative 20; Mismatches 23; Indels 2; Gaps 2;

Db 511 fisekfngdsalrfeqstntarytlrigngsnylyrvssignstirvtgrvvtas 569  
 QY 12 YVOTPFLESNSVRKYIS-IGSCLSTAGPSYVKF-QDNVPGSQTFSAGLHLRVEDPS 68

## RESULT 14

ID R14618 standard; Protein: 1693 AA.  
 AC R14618;  
 DT 16-JAN-1992 (first entry)  
 DE Protein encoded by ORF 1 of Burmese ET-NANB viral strain.  
 KW Enterically transmitted non-A, non-B hepatitis virus; hepatitis C;  
 KW HCV; E.coli strain BB4; ATCC deposit number 67717; Burma.  
 OS Enterically transmitted non-A, non-B hepatitis virus.  
 PN WO9115603-A.  
 PD 17-OCT-1991.  
 PF 05-APR-1991; U02368.  
 PR 05-APR-1990; US-505888.  
 PA (GENE-) GENELABS INC.  
 PA (USSH ) US DEPT HEALTH & HUMAN.  
 PI Reyes GR, Yarbough PO, Bradley DW, Krawczynski KZ, Tam A;  
 PI Fry KE;  
 DR WPI: 91-325242/44.  
 DR N-PSDB: Q14412.  
 DT New viral proteins from non A-non-B hepatitis agent - used to  
 PT treat and prevent enterically-transmitted non-A non-B hepatitis  
 PT virus  
 PS Disclosure; Page 15; 117pp; English.  
 CC A positive clone ET1.1 was identified in a library prepared from  
 CC bile of cynomolgus monkeys infected with the Burma strain of ET-NANB.  
 CC Both strands of ET1.1 were sequenced. Identity of the sequence with  
 CC sequences in etiologic agents has been confirmed by locating a  
 CC similar sequence in a viral strain isolated in Burma. This protein  
 CC is encoded by the longest ORF (ORF 1) of the Burma strain.  
 CC (See Q14410 for ET1.1).  
 SQ Sequence 1693 AA;

Query Match 13.2%; Score 84; DB 3; Length 1693;  
 Best Local Similarity 26.6%; Pred. No. 2.55e+01;  
 Matches 17; Conservative 20; Mismatches 24; Indels 3; Gaps 3;



Search completed: Wed Sep 1 16:32:18 1999  
Job time : 34 secs.

RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,  
RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RL NATURE 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA MILLER N.,  
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA WATERSTON R.,  
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U40951; G107232; -  
SQ SEQUENCE 53 AA; 5833 MW; CADA122E CRC32;  
  
Query Match 64.7%; Score 44; DB 5; Length 53;  
Best Local Similarity 58.3%; Pred. No. 2.09e+01;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
Db 6 STSNSTSRVSFV 17  
||||| : :||  
QY 81 STSNDTTSAAVF 92  
  
RESULT 11  
ID Q00090 PRELIMINARY; PRT; 236 AA.  
AC Q00090;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE POTASSIUM CHANNEL PROTEIN (FRAGMENT).  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA RUDY B., KENTROS C., DE MIERA E.,  
RL MOL. CELL. NEUROSCI. 2:89-102(1991).  
CC -!- FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT POTASSIUM  
ION PERMEABILITY OF EXCITABLE MEMBRANES.  
CC ASSUMING OPENED OR CLOSED CONFORMATIONS IN RESPONSE TO THE  
VOLTAGE DIFFERENCE ACROSS THE MEMBRANE, THE PROTEIN FORMS A  
POTASSIUM-SELECTIVE CHANNEL THROUGH WHICH K+ IONS MAY PASS IN  
ACCORDANCE WITH THEIR ELECTROCHEMICAL GRADIENT.  
CC MAY PLAY A ROLE IN THE NERVOUS SYSTEM AND IN THE REGULATION OF  
BEATING FREQUENCY IN PACEMAKER CELLS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- THE SEGMENT S4 IS PROBABLY THE VOLTAGE-SENSOR AND IS  
CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT  
EVERY THIRD POSITION.  
CC -!- THE AMINO TERMINUS MAY BE IMPORTANT IN DETERMINING THE RATE OF  
INACTIVATION OF THE CHANNEL WHILE THE TAIL MAY PLAY A ROLE IN  
MODULATION OF CHANNEL ACTIVITY AND/OR TARGETING OF THE CHANNEL TO  
SPECIFIC SUBCELLULAR COMPARTMENTS.  
CC -!- SIMILARITY: THIS CHANNEL PROTEIN BELONGS TO THE A-TYPE POTASSIUM  
CURRENT CLASS.  
CC -!- SIMILARITY: BELONGING TO THE SHIV GENE SUBFAMILY, A HOMOLOGUE OF  
THE SHAKER GENE SHAL OF DROSOPHILA MELANOGASTER.  
DR EMBL: M74898; G205041; -  
DR PFAM: PF00520; Ion\_trans; 1.  
KW IONIC CHANNEL; TRANSMEMBRANE; ION TRANSPORT; VOLTAGE-GATED CHANNEL;  
KW MULTIGENE FAMILY.  
FT NON\_TER 1  
FT TRANSMEM 15 33 SEGMENT S1 (BY SIMILARITY).  
FT TRANSMEM 57 78 SEGMENT S2 (BY SIMILARITY).  
FT TRANSMEM 89 110 SEGMENT S3 (BY SIMILARITY).  
FT TRANSMEM 118 136 SEGMENT S4 (BY SIMILARITY).

FT TRANSMEM 152 173 SEGMENT S5 (BY SIMILARITY).  
FT TRANSMEM 213 234 SEGMENT S6 (BY SIMILARITY).  
FT NON\_TER 236  
SQ SEQUENCE 236 AA; 26261 MW; C4A458FB CRC32;  
  
Query Match 64.7%; Score 44; DB 11; Length 236;  
Best Local Similarity 50.0%; Pred. No. 2.09e+01;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
  
Db 108 TNNEDVSGAFVT 119  
|:|:|:|:|:|:  
QY 82 TSNDTTSAAVFS 93  
  
RESULT 12  
ID O65584 PRELIMINARY; PRT; 371 AA.  
AC O65584;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 40.8 KD PROTEIN.  
GN M3E9.70.  
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;  
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BEVAN M., VANDENBOL M., JALLET C., PORTETELLE D., HOHEISEL J.,  
RA MEWES H.W., MAYER K., SCHUELLER C.,  
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU ARABIDOPSIS SEQUENCING PROJECT;  
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AL022223; E1283943; -  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 371 AA; 40833 MW; 580F5AA1 CRC32;  
  
Query Match 64.7%; Score 44; DB 10; Length 371;  
Best Local Similarity 38.5%; Pred. No. 2.09e+01;  
Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;  
  
Db 224 SSSGESSESSFSVS 236  
|:|:|:|:|:|:  
QY 81 STSNDTTSAAVFS 93  
  
RESULT 13  
ID O68923 PRELIMINARY; PRT; 461 AA.  
AC O68923;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE HOMOLOG.  
GN GAPX.  
OS STREPTOMYCES ROSEOFULVUS.  
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;  
OC ACTINOMYCETALES; STREPTOMICINAE; STREPTOMYCETACEAE; STREPTOMYCES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94237486.  
RA BIBB M.J., SHERMAN D.H., OMURA S., HOPWOOD D.A.,  
RT "Cloning, sequencing and deduced functions of a cluster of  
Streptomyces genes probably encoding biosynthesis of the polyketide  
antibiotic frenolicin.";  
RL GENE 142:31-39(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA REEVES C.D., SOLIDAY C.L.,  
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AF058302; G3170587; -  
SQ SEQUENCE 461 AA; 49966 MW; 26EC2672 CRC32;

```
OC EUKARYOTA: DICTYOSTELIIDA: DICTYOSTELIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AX4:
RA LOOMIS W.F.:
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBDJ DATA BANKS.
DR EMBL: U66911; G1519534; -.
FT NON_TER 268 268
SQ SEQUENCE 268 AA; 30503 MW; 04E723F5 CRC32;

Query Match 67.6%; Score 46; DB 5; Length 268;
Best Local Similarity 53.8%; Pred. No. 7.24e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 15 TTNNNTTPEKVS 27
:|:|:|:|
Qy 81 STSNDTTSAAVF 93

RESULT 7
ID Q19474 PRELIMINARY; PRT; 337 AA.
AC Q19474:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE F15A2.4 PROTEIN.
GN F15A2.4.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA: METAZOA: NEMATODA: SECERNENTEA: RHABDITIA: RHABDITIDA:
OC RHABDITINA: RHABDITOIDEA: RHABDITIIDAE: PELODERINAE: CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA GREGORY J.:
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBDJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KESHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
DR EMBL: Z70207; E1345607; -.
SQ SEQUENCE 337 AA; 38483 MW; ABD004B8 CRC32;

Query Match 66.2%; Score 45; DB 5; Length 337;
Best Local Similarity 66.7%; Pred. No. 1.24e+01;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 317 SKHNSSTTSATFV 328
| | | | | | | |
Qy 81 STSNDTTSAAVF 92

RESULT 8
ID Q12372 PRELIMINARY; PRT; 583 AA.
AC Q12372:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CHROMOSOME XII READING FRAME ORF YLLO61W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA: FUNGI: ASCOMYCOTA: HEMIASCOCYCETES: SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE: SACCHAROMYCETES.
RN [1]

RP SEQUENCE FROM N.A.
RA WEDLER H., WEDLER E., SCHARFE M., WAMBUITT R.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBDJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA MIPS;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBDJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN-S288C;
RA WEDLER H., WAMBUITT R.;
RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBDJ DATA BANKS.
DR EMBL: Z73166; E245763; -.
DR EMBL: Z47973; G642323; -.
DR PFAM: PF00324; aa_permeases; 1.
SQ SEQUENCE 583 AA; 64217 MW; BDE5153C CRC32;

Query Match 66.2%; Score 45; DB 3; Length 583;
Best Local Similarity 50.0%; Pred. No. 1.24e+01;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 330 SSEDITASPFV 341
|:|:| |::|
Qy 81 STSNDTTSAAVF 92

RESULT 9
ID Q26818 PRELIMINARY; PRT; 37 AA.
AC Q26818:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SURFACE ANTIGEN (FRAGMENT).
OS TRYPAOSOMA BRUCEI BRUCEI.
OC EUKARYOTA: EULENZOZA: KINETOPLASTIDA: TRYPAOSOMATIDAE: TRYPAOSOMA.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-GAMBIENSE;
RA MERRITT S.C., TSCHUDI C., KONIGSBERG W.H., RICHARDS F.F.;
RT "Reverse transcription of trypanosome variable antigen mRNAs
RT initiated by a specific oligonucleotide primer.";
RL PROC. NATL. ACAD. SCI. U.S.A. 80:1536-1540(1983).
DR EMBL: V01391; G10560; -.
FT NON_TER 1
SQ SEQUENCE 37 AA; 4068 MW; 551442D4 CRC32;

Query Match 64.7%; Score 44; DB 5; Length 37;
Best Local Similarity 63.6%; Pred. No. 2.09e+01;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 14 TSNTTASNSFV 24
| | | | | | | |
Qy 82 TSNDTTSAAVF 92

RESULT 10
ID Q23582 PRELIMINARY; PRT; 53 AA.
AC Q23582:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO ADENOSINE A3 RECEPTOR. NCBI GI: 1072232.
GN ZK721.5.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA: METAZOA: NEMATODA: SECERNENTEA: RHABDITIA: RHABDITIDA:
OC RHABDITINA: RHABDITOIDEA: RHABDITIIDAE: PELODERINAE: CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
```



```
RP SEQUENCE FROM N.A.
RA POWERS P.A., GREGG R.G., HOGAN K.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U86953; G2155256; JOINED.
DR EMBL; U86954; G2155256; JOINED.
DR EMBL; U86955; G2155256; JOINED.
DR EMBL; U86961; G2155256; JOINED.
DR EMBL; U86952; G2155256; JOINED.
DR EMBL; U86956; G2155256; JOINED.
DR EMBL; U86959; G2155256; JOINED.
DR EMBL; U86960; G2155256; JOINED.
DR EMBL; U86957; G2155256; JOINED.
DR EMBL; U86958; G2155256; JOINED.
DR PFAM; PF0018; SH3; 1.
SQ SEQUENCE 579 AA; 63705 MW; 651614EF CRC32;

Query Match 79.4%; Score 54; DB 4; Length 579;
Best Local Similarity 75.0%; Pred. No. 7.52e+02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNFV 55
   |||:||||:|
Qy 81 STSNDTTSAAVF 92

RESULT 3
ID Q20870 PRELIMINARY; PRT; 241 AA.
AC Q20870;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO GLUTAMATE RECEPTORS.
GN F56D1.7.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDIIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPESE T., COOPER J., FAVELLO A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA CHISSOE S., WILSON R.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U39997; G1055122; -.
SQ SEQUENCE 241 AA; 27445 MW; F0977197 CRC32;

Query Match 69.1%; Score 47; DB 5; Length 241;
Best Local Similarity 80.0%; Pred. No. 4.20e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 59 GNDTISAADFV 68
   :||| |||||
Qy 83 SNDTTSAAVF 92

RESULT 4
ID Q09532 PRELIMINARY; PRT; 982 AA.
AC Q09532;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHEICAL 110.9 KD PROTEIN F07F6.6 IN CHROMOSOME II.
GN F07F6.6.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDIIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA CHISSOE S.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: TO GLUTAMATE RECEPTOR.
DR EMBL; U23486; G746452; -.
DR WORMPEP; F07F6.6; CE01898.
DR PFAM; PF00060; lig_chan; 1.
KW HYPOTHEICAL PROTEIN.
FT DOMAIN 827 830 POLY-ARG.
SQ SEQUENCE 982 AA; 110892 MW; B4285770 CRC32;

Query Match 69.1%; Score 47; DB 5; Length 982;
Best Local Similarity 80.0%; Pred. No. 4.20e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 59 GNDTISAADFV 68
   :||| |||||
Qy 83 SNDTTSAAVF 92

RESULT 5
ID P87112 PRELIMINARY; PRT; 2100 AA.
AC P87112;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE PUTATIVE TRANSCRIPTIONAL REGULATOR C20G8.06.
GN SPAC20G8.06.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
RN [1]
RP SEQUENCE FROM N.A.
RA BADCOCK K., CHURCHER C.M., WOOD V., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: TO YEAST NOT1.
DR EMBL; Z95334; E315491; -.
KW HYPOTHEICAL PROTEIN; NUCLEAR PROTEIN; TRANSCRIPTION REGULATION.
SQ SEQUENCE 2100 AA; 237132 MW; 9A710DB4 CRC32;

Query Match 69.1%; Score 47; DB 3; Length 2100;
Best Local Similarity 66.7%; Pred. No. 4.20e+00;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1601 SATNDTAAVF 1612
   :||| |||||
Qy 81 STSNDTTSAAVF 92

RESULT 6
ID Q94484 PRELIMINARY; PRT; 268 AA.
AC Q94484;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)
DE ORF DG1007 (FRAGMENT).
OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
```

(TM)

Result No.	Score	Query		ID	Description	Pred. No.	
		Match	Length				
1	54	79.4	101	11	0885517	CALCIUM CHANNEL BETA 1	7.52e-02
2	54	79.4	579	4	015331	L-TYPE CALCIUM CHANNEL	7.52e-02
3	47	69.1	241	5	Q20870	SIMILAR TO GLUTAMATE R	4.20e+00
4	47	69.1	982	5	P09512	HYPOTHETICAL 110.9 KD	4.20e+00
5	47	69.1	2100	3	Q97132	PUTATIVE TRANSCRIPTION	4.20e+00
6	46	67.6	268	5	Q94484	ORF DG1007 (FRAGMENT).	4.20e+00
7	45	66.2	337	5	Q19474	ORF15A2.4 PROTEIN.	7.24e+00
8	45	66.2	583	3	Q23732	CHROMOSOME XII READING	1.24e+01
9	44	64.7	37	5	Q26818	SURFACE ANTIGEN (FRAG	2.09e+01
10	44	64.7	53	5	Q23582	SIMILAR TO ADENOSINE A	2.09e+01
11	44	64.7	236	11	Q00090	POTASSIUM CHANNEL PROT	2.09e+01
12	44	64.7	371	10	Q85584	HYPOTHETICAL 40.8 KD P	2.09e+01
13	44	64.7	461	2	Q88923	GLYCERALDEHYDE-3-PHOS	2.09e+01
14	44	64.7	488	5	Q43317	REVERSE TRANSCRIPTASE	2.09e+01
15	44	64.7	545	3	P87306	HYPOTHETICAL 62.5 KD P	2.09e+01
16	44	64.7	575	5	Q26040	SHAL 1 POTASSIUM CHANN	2.09e+01
17	44	64.7	611	11	Q63286	VOLTAGE-GATED K+ CHANN	2.09e+01
18	44	64.7	623	5	Q16932	W3F9.4 PROTEIN.	2.09e+01
19	44	64.7	636	11	P70632	KV4.3 POTASSIUM CHANNE	2.09e+01
20	44	64.7	636	4	Q60576	KV4.3 POTASSIUM CHANNE	2.09e+01

```

CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; U67505; G1591267;
CC      TIGR; MJ0561;
CC      PROSITE; PS01266; ADENYLOSUCCIN_SYN_1; 1.
CC      PFAM; PF00709; ADENYLOSUCCIN_SYN_2; FALSE_NEG.
CC      HSSP; P12283; IADI.
CC      PURINE BIOSYNTHESIS; LIGASE; GTP-BINDING.
CC      NP_BIND 18 24
CC      SEQUENCE 345 AA; 37820 MW; E15BEZC9 CRC32;
CC
Query Match      66.2%; Score 45; DB 1; Length 345;
Best Local Similarity 50.0%; Pred. No. 5.97e+00;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 200 TSKDNTASSPRA 211
QY 82 TSDNTTSAAFVS 93
  |||:||||:
  -----
RESULT 14
ID IR12_HCMVA STANDARD; PRT; 416 AA.
AC P16810;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DE 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN IRL12 (TFL12).
OS HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
CC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
CC BETAHERPESVIRINAE; CYTOMEGALOVIRUS.
CC
SEQUENCE FROM N.A.
RX MEDLINE; 90269039.
RA CHEE M.S., BANKIER A.T., BECK S., BOHNI R., BROWN C.M., CERNY R.,
RA HORNELL T., HUTCHISON C.A. III, KOUZARIDES T., MARTIGNETTI J.A.,
RA PREDDIE E., SATCHWELL S.C., TOMLINSON P., WESTON K.M., BARRELL B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169."
RL CURR. TOP. MICROBIOL. IMMUNOL. 154:125-169(1990).
CC
-1- SIMILARITY: BELONGS TO RL11 FAMILY.
CC
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
-----
CC      EMBL; X17403; E298604;
CC      EMBL; X17403; E298631;
CC      PIR; S09761; S09761.
CC      HYPOTHETICAL PROTEIN.
CC      SEQUENCE 416 AA; 47416 MW; 8289EB89 CRC32;
CC
Query Match      66.2%; Score 45; DB 1; Length 416;
Best Local Similarity 50.0%; Pred. No. 5.97e+00;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 42 TSPNTASTTFTV 53
QY 82 TSDNTTSAAFVS 93
  |||:||||:
  -----
RESULT 15
ID POLG_BCMVN STANDARD; PRT; 3066 AA.
AC Q85399;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER
DE COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KD PROTEIN

```

```

DE 1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KD PROTEIN 2 (6K2);
DE GENOME-LINKED PROTEIN (VPC); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)
DE (EC 3.4.22.-) (49 KD PROTEINASE) (49 KD-PRO); NUCLEAR INCLUSION
DE PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);
DE COAT PROTEIN (CP)].
OS BEAN COMMON MOSAIC VIRUS (STRAIN NL-3 / MICHIGAN) (BCMV).
CC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
CC POTYVIRUS.
CC
SEQUENCE FROM N.A.
RX MEDLINE; 96191623.
RA FANG G.W., ALLISON R.F., ZAMBOLIM E.M., MAXWELL D.P., GILBERTSON R.L.;
RT "The complete nucleotide sequence and genome organization of bean
RT common mosaic virus (NL3 strain).";
RL VIRUS RES. 39:13-23(1995).
CC
-1- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID
CC TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
CC
-1- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
CC MAY BE INVOLVED IN REPLICATION.
CC
-1- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
CC
-1- PTM: VPC IS COVALENTLY LINKED TO THE GENOMIC RNA.
CC
-1- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
CC POLYPEPTIDE WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
CC INDIVIDUAL PROTEINS.
CC
-1- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
CC
-1- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
CC
-1- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.
CC
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
-----
CC      EMBL; U19287; G1373002;
CC      PFAM; PF00271; helicase_C; 1.
CC      PFAM; PF00680; RNA_dep_RNA_pol; 1.
CC      PFAM; PF00767; Poty_coat; 1.
CC      PFAM; PF00851; Peptidase_C6; 1.
CC      PFAM; PF00863; Peptidase_C4; 1.
CC      HYDROLASE; TRANSFERASE; THIOL PROTEASE; RNA-DIRECTED RNA POLYMERASE;
CC      COAT PROTEIN; POLYPROTEIN; COVALENT PROTEIN-RNA LINKAGE; HELICASE;
CC      ATP-BINDING.
CC      CHAIN 1 ? ? N-TERMINAL PROTEIN.
CC      CHAIN ? ? ? HELPER COMPONENT PROTEINASE.
CC      CHAIN ? ? ? PROTEIN P3.
CC      CHAIN ? ? ? 6 KD PROTEIN 1.
CC      CHAIN ? ? ? CYTOPLASMIC INCLUSION PROTEIN.
CC      CHAIN ? ? ? 6 KD PROTEIN 2.
CC      CHAIN ? ? ? GENOME-LINKED PROTEIN.
CC      CHAIN ? ? ? NUCLEAR INCLUSION PROTEIN A.
CC      CHAIN ? ? ? NUCLEAR INCLUSION PROTEIN B.
CC      CHAIN ? 3066 COAT PROTEIN.
CC      NP_BIND 1258 1265 ATP (POTENTIAL)
CC      SEQUENCE 3066 AA; 350387 MW; 67B875D6 CRC32;
CC
Query Match      66.2%; Score 45; DB 1; Length 3066;
Best Local Similarity 54.5%; Pred. No. 5.97e+00;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 1194 STKESSTSTSF 1204
QY 81 TSDNTTSAAF 91
  |||:||||:
  -----
Search completed: Wed Sep 1 16:31:25 1999
Job time : 11 secs.

```

SQ SEQUENCE 1036 AA: 109310 MW: CCDE5BE8 CRC32;  
 Query Match 70.6%; Score 48; DB 1; Length 1036;  
 Best Local Similarity 53.8%; Pred. No. 1.11e+00;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 87 SSSNNNTSNGFVA 99  
 I::I::I::I::I::  
 QY 81 STSNDTTSAAAFVS 93

RESULT 11  
 ID CYP1DORAC STANDARD; PRT; 532 AA.  
 AC Q44220;  
 DT 15-DEC-1998 (REL. 37, CREATED)  
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE CYTOCHROME P450 12B1 PRECURSOR (EC 1.14.-.-).  
 GN CYP12B1.  
 OS DROSOPHILA ACANTHOPTERA (FRUIT FLY).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;  
 OC DROSOPHILIDAE; DROSOPHILA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A584.2;  
 RX MEDLINE; 97449781.  
 RA DANIELSON P.B., FOGLEMAN J.C.;  
 RA "Isolation and sequence analysis of cytochrome P450 12B1: the first  
 RT mitochondrial insect P450 with homology to 1 alpha,25 dihydroxy-D3  
 RT 24-hydroxylase.";  
 RL INSECT BIOCHEM. MOL. BIOL. 27:595-604(1997).  
 CC -1- FUNCTION: PROBABLY INVOLVED IN STEROID HORMONES BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U78485; G2674278; -  
 CC HYPOTHETICAL PROTEIN.  
 DR QUERY MATCH 67.6%; Score 46; DB 1; Length 532;  
 SQ SEQUENCE 532 AA: 62031 MW: 53F97355 CRC32;  
 Query Match 67.6%; Score 46; DB 1; Length 532;  
 Best Local Similarity 77.8%; Pred. No. 3.44e+00;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 335 DTTSSAFVT 343  
 I::I::I::I::I::  
 QY 85 DTTSSAAAFVS 93

RESULT 12  
 ID YEM3 YEAST STANDARD; PRT; 1076 AA.  
 AC P40021;  
 DT 01-FEB-1995 (REL. 31, CREATED)  
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 119.3 K D PROTEIN IN PIP1-GLN3 INTERGENIC REGION.  
 GN YER033C.  
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;

OC SACCHAROMYCETACEAE; SACCHAROMYCES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA DIETRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,  
 RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,  
 RA CHUNG E., DUNCAN M., GUZMAN E., HARTZELL G., HUNICKE-SMITH S.,  
 RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,  
 RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGREN R., OEFNER P., OH C.,  
 RA PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOOREN T., SMITH V.,  
 RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.;  
 RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U18796; G603266; -  
 CC HYPOTHETICAL PROTEIN.  
 DR QUERY MATCH 67.6%; Score 46; DB 1; Length 1076;  
 SQ SEQUENCE 1076 AA: 119349 MW: 9FED783D CRC32;  
 Query Match 67.6%; Score 46; DB 1; Length 1076;  
 Best Local Similarity 50.0%; Pred. No. 3.44e+00;  
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 97 TSTNQTTTSNFV 108  
 I::I::I::I::I::  
 QY 81 STSNDTTSAAAFV 92

RESULT 13  
 ID PURA\_MERJA STANDARD; PRT; 345 AA.  
 AC Q57981;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE ADENYLOSUCCLINATE SYNTHETASE (EC 6.3.4.4) (IMP--ASPARTATE LIGASE).  
 GN PURA OR MJ0561.  
 OS METHANOCOCCUS JANNASCHII.  
 OC ARCHAEA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;  
 OC METHANOCOCCUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE; 96337999.  
 RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,  
 RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,  
 RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,  
 RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,  
 RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,  
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,  
 RA KLENK H.-P., FRASER C.M., SMITH H.O., WOENSE C.R., VENTER J.C.;  
 RL "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RT jannaschii.";  
 RT SCIENCE 273:1058-1073(1996).  
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE DE NOVO PATHWAY OF PURINE  
 CC NUCLEOTIDE BIOSYNTHESIS.  
 CC -1- CATALYTIC ACTIVITY: GTP + IMP + L-ASPARTATE -> GDP +  
 CC ORTHOPHOSPHATE + ADENYLOSUCCLINATE.  
 CC -1- PATHWAY: FIRST COMMITTED STEP IN AMP BIOSYNTHESIS.  
 CC -1- SIMILARITY: WITH OTHER ADENYLOSUCCLINATE SYNTHETASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>



```

RESULT 7
ID CCB2_RABIT STANDARD; PRT; 632 AA.
AC P54298;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-2 SUBUNIT
DE (CAB2).
GN CACNB2 OR CACNLB2.
OS ORYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RX MEDLINE; 92192022.
RA HULLIN R., SINGER-LAHAT D., FREICHEL M., BIEL M., DASCAL N.,
RA HOFMANN F., FLOCKERZI V.;
RT "Calcium channel beta subunit heterogeneity: functional expression of
RT cloned cDNA from heart, aorta and brain.";
RL EMBO J. 11:885-890(1992)
CC -1- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)
CC SENSITIVE CALCIUM CHANNEL.
CC -1- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:
CC ALPHA-1, ALPHA-2, BETA AND GAMMA.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN HEART, AORTA AND
CC BRAIN.
CC -1- ALTERNATIVE PRODUCTS: THREE FORMS OF THE BETA-2 SUBUNIT ARE
CC PRODUCED BY ALTERNATIVE SPLICING OF THE GENE: CAB2A, CAB2B (SHOWN
CC HERE) AND CAB2C.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64298; G1500; -
DR EMBL; X64297; G1498; -
DR EMBL; X64299; G1502; -
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00774; Ca_channel_B; 1.
KW IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;
KW GLYCOPROTEIN; PHOSPHORYLATION; SH3 DOMAIN; MULTIGENE FAMILY;
KW ALTERNATIVE SPLICING.
FT DOMAIN 86 147 SH3.
FT CARBOHYD 94 94 POTENTIAL.
FT CARBOHYD 175 175 POTENTIAL.
FT VARSPPLIC 1 43
FT FT
FT VARSPPLIC 197 234
FT VARSPPLIC 235 240 MISSING (IN CAB2C).
FT MPEFKK -> KQKQKS (IN CAB2C).
SQ SEQUENCE 632 AA; 70943 MW; DF97DA5D CRC32;

Query Match 79.4%; Score 54; DB 1; Length 632;
Best Local Similarity 75.0%; Pred. No. 3.03e-02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 30 STSDTTSNSFV 41
| | | | | | | |
QY 81 STSDTTSAAFY 92

RESULT 8
ID LPFD_SALTY STANDARD; PRT; 355 AA.
AC P43663;

01-NOV-1995 (REL. 32, CREATED)
01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
LPFD PROTEIN PRECURSOR.
GN LPFD.
OS SALMONELLA TYPHIMURIUM.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 14028;
RX MEDLINE; 95238281.
RA BAUEMLER A.J., HEFFRON F.;
RT "Identification and sequence analysis of ipfABCDE, a putative
RT fibrial operon of Salmonella typhimurium.";
RL J. BACTERIOL. 177:2087-2097(1995).
CC -1- SIMILARITY: BELONGS TO THE FMH / LPFD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U18559; G829374; -
DR STYGENE; SG10560; LPFD.
DR PFAM; PF00419; Fimbrial; 1.
KW FIMBRIA; SIGNAL.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 355 LPFD PROTEIN.
SQ SEQUENCE 355 AA; 37714 MW; 20D529B5 CRC32;

Query Match 70.6%; Score 48; DB 1; Length 355;
Best Local Similarity 63.6%; Pred. No. 1.11e+00;
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 51 TSKNTTCAFTV 61
| | | | | | | |
QY 82 TSDNTTSAAFV 92

RESULT 9
ID EMR1_HUMAN STANDARD; PRT; 886 AA.
AC Q14246;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CELL SURFACE GLYCOPROTEIN EMR1 PRECURSOR (EMR1 HORMONE RECEPTOR).
GN EMR1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95324926.
RA BAUD V., CHISSOE S.L., VIEGAS-PEQUIGNOT E., DIRIONG S., N'GUYEN V.C.,
RA ROE B.A., LIPINSKI M.;
RT "EMR1, an unusual member in the family of hormone receptors with
RT seven transmembrane segments.";
RL GENOMICS 26:334-344(1995).
CC -1- FUNCTION: PROBABLY INVOLVED IN CELLULAR RESPONSE TO A HORMONE OR
CC AN INTERACTION WITH A PROTEIN LIGAND.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: WIDE EXPRESSION; INCREASED LEVELS IN
CC PERIPHERAL BLOOD MONONUCLEAR CELLS.
CC -1- PTM: N- AND O-GLYCOSYLATED; (POSSIBLE).
CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

DIHYDROPYRIDINE-SENSITIVE L-TYPE, CHANNEL BETA-1-B2 SUBUNIT  
(BETA-1 ISOFORM A).  
GN CACNB1 OR CACNLB1.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-HIPOCAMPUS;  
RX POWERS P.A., LIU S., HOGAN K., GREGG R.G.;  
RT "Skeletal muscle and brain isoforms of a beta-subunit of human  
voltage-dependent calcium channels are encoded by a single gene.";  
RL J. BIOL. CHEM. 267:22967-22972(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-HEART;  
RX MEDLINE; 93265672.  
RA COLLIN T., WANG J., NARGEOT J., SCHWARTZ A.;  
RT "Molecular cloning of three isoforms of the L-type voltage-dependent  
calcium channel beta subunit from normal human heart.";  
RL CIRC. RES. 72:1337-1344(1993).  
CC -!- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)  
SENSITIVE CALCIUM CHANNEL.  
CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:  
ALPHA-1, ALPHA-2, BETA AND GAMMA.  
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH  
THE CYTOPLASMIC ASPECT OF THE SARCOLEMMMA.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, HEART AND SPLEEN.  
CC -!- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1  
SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE.  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M92303; G179806; -  
DR EMBL; L06110; G187015; -  
DR PIR; C44461; C44461.  
DR MIN; I14207; -  
DR PROSITE; PS50002; SH3; 1.  
DR PFAM; PF00018; SH3; 1.  
DR PFAM; PF00774; Ca\_channel\_B; 1.  
KW IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;  
KW GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;  
KW MULTIGENE FAMILY.  
FT DOMAIN 100 161  
FT CARBOHYD 189 189  
FT CARBOHYD 425 425  
FT POTENTIAL.  
FT TS -> SG (IN REF. 2).  
FT G -> R (IN REF. 2).  
FT H -> D (IN REF. 2).  
FT KL -> TV (IN REF. 2).  
FT G -> S (IN REF. 2).  
FT T -> S (IN REF. 2).  
FT SNTR -> LQHT (IN REF. 2).  
FT I -> L (IN REF. 2).  
FT M -> I (IN REF. 2).  
FT AA -> RR (IN REF. 1).  
FT E -> D (IN REF. 2).  
FT M -> V (IN REF. 2).  
FT R -> W (IN REF. 2).  
FT L -> P (IN REF. 2).  
FT L -> P (IN REF. 2).  
FT A -> GTP (IN REF. 2).  
FT MISSING (IN REF. 2).

FT CONFLICT 557 557 L -> M (IN REF. 2).  
FT CONFLICT 571 572 WP -> CA (IN REF. 2).  
FT CONFLICT 591 591 R -> Q (IN REF. 2).  
SQ SEQUENCE 596 AA; 65638 MW; OCC447D4 CRC32;  
Query Match 79.4%; Score 54; DB 1; Length 596;  
Best Local Similarity 75.0%; Pred. No. 3.03e-02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Db 44 STSDTTTNSFV 55  
Qy 81 STSNDTTSAFV 92  
|||||:||||:|  
RESULT 6  
ID CCBA\_RAT STANDARD; PRT; 597 AA.  
AC P54283;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CHANNEL BETA-1-B2 SUBUNIT  
(BETA-1 ISOFORM A).  
DE CACNB1 OR CACNLB1.  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RX MEDLINE; 92038046.  
RA PRAGNELL M., SAKAMOTO J., JAY S.D., CAMPBELL K.P.;  
RT "Cloning and tissue-specific expression of the brain calcium channel  
beta-subunit";  
RL FEBS LETT. 291:253-258(1991).  
CC -!- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)  
SENSITIVE CALCIUM CHANNEL.  
CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:  
ALPHA-1, ALPHA-2, BETA AND GAMMA.  
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH  
THE CYTOPLASMIC ASPECT OF THE SARCOLEMMMA.  
CC -!- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1  
SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE (BY  
SIMILARITY).  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X61394; G55894; -  
DR PROSITE; PS50002; SH3; 1.  
DR PFAM; PF00018; SH3; 1.  
DR PFAM; PF00774; Ca\_channel\_B; 1.  
KW IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;  
KW GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;  
KW MULTIGENE FAMILY.  
FT DOMAIN 100 161  
FT CARBOHYD 189 189  
FT CARBOHYD 425 425  
FT POTENTIAL.  
SQ SEQUENCE 597 AA; 65679 MW; 23B08C47 CRC32;  
Query Match 79.4%; Score 54; DB 1; Length 597;  
Best Local Similarity 75.0%; Pred. No. 3.03e-02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Db 44 STSDTTTNSFV 55  
Qy 81 STSNDTTSAFV 92  
|||||:||||:|

Db 44 STSDTTSNSFV 55  
|||:||||:|  
Qy 81 STSNDTTSAAVF 92

RESULT 3  
ID CCB2\_HUMAN STANDARD; PRT: 524 AA.  
AC P19517;  
DT 01-FEB-1991 (REL. 17, CREATED)  
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-1M SUBUNIT  
DE (BETA-1 ISOFORM C).  
GN CACNB1 OR CACNLB1.  
OS ORCTOLOGUS CUNICULUS (RABBIT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC LAGOMORPHA; LEPOPIDAE; ORCTOLOGUS.  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE-SKELETAL MUSCLE;  
RX MEDLINE; 89368946.  
RA FLOCKERZI V., HOFFMANN F.;  
RA "Primary structure of the beta subunit of the DHP-sensitive calcium  
RT channel from skeletal muscle.";  
RL SCIENCE 245:1115-1118(1989).  
CC -!- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)  
CC SENSITIVE CALCIUM CHANNEL.  
CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:  
CC ALPHA-1, ALPHA-2, BETA AND GAMMA.  
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH  
CC THE CYTOPLASMIC ASPECT OF THE SARCOLEMA.  
CC -!- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1  
CC SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE (BY  
CC SIMILARITY).  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; M25817; G164834; -  
CC EMBL; M25514; G598370; -  
CC PROSITE; PS50002; SH3; 1.  
CC PFAM; PF00018; SH3; 1.  
CC PFAM; PF00774; Ca-channel\_B; 1.  
CC IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;  
CC GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;  
CC MULTIGENE FAMILY.  
FT MOD\_RES 71 71 BLOCKED.  
FT DOMAIN 100 161 SH3.  
FT CARBOHYD 189 189 POTENTIAL.  
FT CARBOHYD 470 470 POTENTIAL.  
FT CARBOHYD 499 499 POTENTIAL.  
FT CONFLICT 304 304 L -> R (IN G598370).  
SQ SEQUENCE 524 AA; 57825 MW; 8AAC8B67 CRC32;  
Query Match 79.4%; Score 54; DB 1; Length 524;  
Best Local Similarity 75.0%; Pred. No. 3.03e-02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNSFV 55  
|||:||||:|  
Qy 81 STSNDTTSAAVF 92

RESULT 4  
ID CCB2\_HUMAN STANDARD; PRT: 567 AA.  
AC Q08289;  
DT 01-OCT-1994 (REL. 30, CREATED)  
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-2 SUBUNIT  
DE (LAMBERT-EATON MYASTHENIC SYNDROME ANTIGEN B) (MYSB).  
GN CACNB2 OR CACNLB2 OR MYSB.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-FETAL BRAIN;  
RX MEDLINE; 93263585.  
RA ROSENFELD M.R., WONG E., DALMAU J., MANLEY G., POSNER J.B.,  
RA SHER E., FURNEAUX H.M.;  
RT "Cloning and characterization of a Lambert-Eaton myasthenic syndrome  
RT antigen.";  
RL ANN. NEUROL. 33:113-120(1993).  
CC -!- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)  
CC SENSITIVE CALCIUM CHANNEL (BY SIMILARITY).  
CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:  
CC ALPHA-1, ALPHA-2, BETA AND GAMMA.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES.  
CC -!- ALTERNATIVE PRODUCTS: THREE FORMS ARE PRODUCED BY ALTERNATIVE  
CC SPLICING OF THE GENE.  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; S60415; G300417; -  
CC PIR; A48895; A48895.  
CC MIN; 600003; -  
CC PROSITE; PS50002; SH3; 1.  
CC PFAM; PF00018; SH3; 1.  
CC PFAM; PF00774; Ca-channel\_B; 1.  
CC IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;  
CC GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;  
CC ANTIGEN.  
FT DOMAIN 114 175 SH3.  
FT CARBOHYD 122 122 POTENTIAL.  
FT CARBOHYD 203 203 POTENTIAL.  
FT CARBOHYD 523 523 POTENTIAL.  
FT VARSPIC 224 268  
FT FT  
FT FT  
FT FT  
FT FT  
FT FT  
FT FT  
SQ SEQUENCE 567 AA; 62087 MW; 4559C166 CRC32;  
Query Match 79.4%; Score 54; DB 1; Length 567;  
Best Local Similarity 75.0%; Pred. No. 3.03e-02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 58 STSDTTSNSFV 69  
|||:||||:|  
Qy 81 STSNDTTSAAVF 92

RESULT 5  
ID CCB2\_HUMAN STANDARD; PRT: 596 AA.  
AC Q02641;  
DT 01-JUN-1994 (REL. 29, CREATED)  
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)



CC -!- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1  
 CC SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE.  
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M92302; G179804; -  
 CC EMBL: M76560; G179744; -  
 CC EMBL: L06111; G187017; -  
 CC EMBL: U86960; G215254; -  
 CC EMBL: U86952; G215254; JOINED.  
 CC EMBL: U86953; G215254; JOINED.  
 CC EMBL: U86954; G215254; JOINED.  
 CC EMBL: U86955; G215254; JOINED.  
 CC EMBL: U86956; G215254; JOINED.  
 CC EMBL: U86957; G215254; JOINED.  
 CC EMBL: U86958; G215254; JOINED.  
 CC EMBL: U86959; G215254; JOINED.  
 CC PIR: B44461; B44461.  
 CC MIM: 114207; -  
 CC PROSITE: PS50002; SH3; 1.  
 CC PFAM: PF00018; SH3; 1.  
 CC PFAM: PF00774; Ca\_channel\_B; 1.  
 CC IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;  
 CC GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;  
 CC MULTIGENE FAMILY.  
 CC DOMAIN 100 161 SH3  
 CC FT CARBOHYD 189 189 POTENTIAL.  
 CC FT CARBOHYD 425 425 POTENTIAL.  
 CC FT CONFLICT 21 21 E -> G (IN REF. 2).  
 CC FT CONFLICT 28 29 QG -> R (IN REF. 3).  
 CC FT CONFLICT 434 435 AA -> RR (IN REF. 1 AND 4).  
 CC SQ SEQUENCE 478 AA; 53006 MW; 6011892D CRC32;  
 CC -----  
 CC Query Match 79.4%; Score 54; DB 1; Length 478;  
 CC Best Local Similarity 75.0%; Pred. No. 3.03e-02;  
 CC Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 CC -----  
 CC Db 44 STSDTTNSFV 55  
 CC QY 81 STSDTTSAFV 92  
 CC III:IIII:II  
 CC -----  
 CC RESULT 2  
 CC ID CCBC\_HUMAN STANDARD; PRT; 523 AA.  
 CC AC Q02639;  
 CC DT 01-JUN-1994 (REL. 29, CREATED)  
 CC DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 CC DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 CC DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-1M SUBUNIT  
 CC DE (BETA-1 ISOFORM C) (BETA-1A).  
 CC GN CACNB1 OR CACNLB1.  
 CC OS HOMO SAPIENS (HUMAN).  
 CC OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 CC OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC TISSUE-SKELETAL MUSCLE;  
 CC RX MEDLINE: 93054616  
 CC RA POWERS P.A., LIO S., HOGAN K., GREGG R.G.;  
 CC RT "Skeletal muscle and brain isoforms of a beta-subunit of human  
 CC RT voltage-dependent calcium channels are encoded by a single gene."  
 CC RL J. BIOL. CHEM. 267:22967-22972(1992).  
 CC RN [2]  
 CC RP SEQUENCE FROM N.A.  
 CC RC TISSUE-HEART;

RX MEDLINE: 93265672.  
 RA COLLIN T., WANG J., NARGEOT J., SCHWARTZ A.;  
 RT "Molecular cloning of three isoforms of the L-type voltage-dependent  
 RT calcium channel beta subunit from normal human heart."  
 RL CIRC. RES. 72:1337-1344(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA POWERS P.A., GREGG R.G., HOGAN K.;  
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [4]  
 RP SEQUENCE OF 146-209 FROM N.A.  
 RX MEDLINE: 93372845.  
 RA ILES D.E., SEGERS B., SENGERS R.C.A., MONSIEURS K., HEYTENS L.,  
 RA HALSALL P.J., HOPKINS P.M., ELLIS F.R., HALL-CURRAN J.L.,  
 RA STEWART A.D., WIERINGA B.;  
 RT "Genetic mapping of the beta 1- and gamma-subunits of the human  
 RT skeletal muscle L-type voltage-dependent calcium channel on  
 RT chromosome 17q and exclusion as candidate genes for malignant  
 RT hyperthermia susceptibility."  
 RL HUM. MOL. GENET. 2:863-868(1993).  
 CC -!- FUNCTION: THIS PROTEIN IS THE BETA SUBUNIT OF THE DIHYDROPYRIDINE  
 CC (DHP) SENSITIVE CALCIUM CHANNEL. PLAYS A ROLE IN EXCITATION-  
 CC CONTRACTION COUPLING. THE SKELETAL MUSCLE DHP-SENSITIVE CA(2+)  
 CC CHANNEL MAY FUNCTION ONLY AS A MULTIPLE SUBUNIT COMPLEX.  
 CC -!- SUBUNIT: THE SKELETAL MUSCLE L-TYPE CALCIUM CHANNEL IS COMPOSED OF  
 CC FIVE SUBUNITS: ALPHA-1, ALPHA-2, BETA, GAMMA AND DELTA.  
 CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH  
 CC THE CYTOPLASMIC ASPECT OF THE SARCOLEMA.  
 CC -!- TISSUE SPECIFICITY: SKELETAL MUSCLE.  
 CC -!- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1  
 CC SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE.  
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M92301; G179802; -  
 CC EMBL: L06112; G187019; -  
 CC EMBL: U86960; G215255; -  
 CC EMBL: U86952; G215255; JOINED.  
 CC EMBL: U86953; G215255; JOINED.  
 CC EMBL: U86954; G215255; JOINED.  
 CC EMBL: U86955; G215255; JOINED.  
 CC EMBL: U86956; G215255; JOINED.  
 CC EMBL: U86957; G215255; JOINED.  
 CC EMBL: U86958; G215255; JOINED.  
 CC EMBL: U86959; G215255; JOINED.  
 CC EMBL: Z21725; G38563; -  
 CC EMBL: Z21726; G38565; -  
 CC MIM: 114207; -  
 CC PROSITE: PS50002; SH3; 1.  
 CC PFAM: PF00018; SH3; 1.  
 CC PFAM: PF00774; Ca\_channel\_B; 1.  
 CC IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;  
 CC GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;  
 CC MULTIGENE FAMILY.  
 CC DOMAIN 100 161 SH3  
 CC FT CARBOHYD 189 189 POTENTIAL.  
 CC FT CARBOHYD 470 470 POTENTIAL.  
 CC FT CONFLICT 28 29 QG -> R (IN REF. 2).  
 CC FT CONFLICT 183 183 MISSING (IN REF. 4).  
 CC FT CONFLICT 479 480 AA -> RR (IN REF. 1 AND 3).  
 CC SQ SEQUENCE 523 AA; 57863 MW; 32CBEE9D CRC32;  
 CC -----  
 CC Query Match 79.4%; Score 54; DB 1; Length 523;  
 CC Best Local Similarity 75.0%; Pred. No. 3.03e-02;  
 CC Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Result No.	Score	Query		Length	DB	ID	Description	Pred. No.
		Match	Score					
1	54	79.4	478	1	CCBB_HUMAN		DIHYDROPYRIDINE-SENSIT	3.03e-02
2	54	79.4	523	1	CCBC_HUMAN		DIHYDROPYRIDINE-SENSIT	3.03e-02
3	54	79.4	524	1	CCBC_RABIT		DIHYDROPYRIDINE-SENSIT	3.03e-02
4	54	79.4	567	1	CCB2_HUMAN		DIHYDROPYRIDINE-SENSIT	3.03e-02
5	54	79.4	596	1	CCBA_HUMAN		DIHYDROPYRIDINE-SENSIT	3.03e-02
6	54	79.4	597	1	CCBA_RAT		DIHYDROPYRIDINE-SENSIT	3.03e-02
7	54	79.4	632	1	CCB2_RABIT		DIHYDROPYRIDINE-SENSIT	3.03e-02
8	48	70.6	355	1	LPFD_SALTY		LPFD PROTEIN PRECURSOR	1.11e+00
9	48	70.6	886	1	EMR1_HUMAN		CELL SURFACE GLYCOPROT	1.11e+00
10	48	70.6	1036	1	NIT2_NEUCR		NITROGEN CATABOLIC ENZ	1.11e+00
11	46	67.6	532	1	CPPL_DROAC		CYTCHROME P450 12B1 P	3.44e+00
12	46	67.6	1076	1	YEM5_YEAST		HYPOTHETICAL I19.3 KD	3.44e+00
13	45	66.2	345	1	PURA_METJA		ADENYLOSUCINATE SYNTH	5.97e+00
14	45	66.2	416	1	IR12_HCMVA		HYPOTHETICAL PROTEIN I	5.97e+00
15	45	66.2	3066	1	POLG_BCMVN		GENOME POLYPROTEIN [CO	5.97e+00
16	44	64.7	490	1	C1KL_DROME		VOLTAGE-GATED POTASSIU	1.03e+01
17	44	64.7	508	1	PR28_YEAST		PRE-mRNA SPLICING FACT	1.03e+01
18	44	64.7	607	1	HRAL_XANCY		HYPERSENSITIVITY RESPO	1.03e+01
19	44	64.7	3106	1	LMA2_MOUSE		LAMININ ALPHA-2 CHAIN	1.03e+01
20	44	64.7	3110	1	LMA2_MOUSE		LAMININ ALPHA-2 CHAIN	1.03e+01
21	43	63.2	142	1	RPC9_YEAST		DNA-DIRECTED RNA POLYM	1.75e+01
22	43	63.2	269	1	FOLD_MYCPN		METHYLENETETRAHYDROFOL	1.75e+01
23	43	63.2	337	1	YEB7_YEAST		HYPOTHETICAL 38.2 KD P	1.75e+01

```
#submission submitted to the EMBL Data Library, December 1994
#description The sequence of S. cerevisiae cosmids 9379, 9581, and lambda
clone 4678.
#accession S50536
##molecule_type DNA
##residues 1-1076 ##label DIE
##cross-references EMBL:U18796; NID:9603265; PID:9603266; MIPS:YER033c
GENETICS
#map_position 5R
SUMMARY #length 1076 #molecular-weight 119349 #checksum 9442
Query Match 67.6%; Score 46; DB 2; Length 1076;
Best Local Similarity 50.0%; Pred. No. 9.97e+00;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 97 TSTNQITTSNFV 108
QY 81 STSNDTISAQV 92
:::||||:|
:::||||:|

Search completed: Wed Sep 1 16:32:49 1999
Job time : 14 secs.
```

```

|||||:|
Qy 81 STSNDTTSAAFV 92

RESULT 11
ENTRY #21048 #type complete
TITLE calcium channel protein beta chain CaB2b - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
DATE 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change
10-Sep-1997

ACCESSIONS S21048
REFERENCE #21046
#authors Hulin, R.; Singer-Lahat, D.; Freichel, M.; Biel, M.; Dascal,
N.; Hofmann, F.; Flockerzi, V.
#journal EMBO J. (1992) 11:885-890
#title Calcium channel beta subunit heterogeneity: functional
expression of cloned cDNA from heart, aorta and brain.
#cross-references MUID:92192022
#accession S21048
#molecule_type mRNA
#residues 1-632 #label HUL
#cross-references EMBL:X64298; NID:g1499; PID:g1500
SUMMARY #length 632 #molecular-weight 70943 #checksum 1583

Query Match 79.4%; Score 54; DB 2; Length 632;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 30 STSDDTNSFV 41
|||||:|
Qy 81 STSNDTTSAAFV 92

RESULT 12
ENTRY #D56271 #type complete
TITLE long polar fimbrial operon protein LpfD - Salmonella
typhimurium
ORGANISM #formal_name Salmonella typhimurium
DATE 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change
09-Sep-1997

ACCESSIONS D56271
REFERENCE A56271
#authors Baumber, A.J.; Heffron, F.
#journal J. Bacteriol. (1995) 177:2087-2097
#title Identification and sequence analysis of lpfABCD, a putative
fimbrial operon of Salmonella typhimurium.
#accession D56271
#status Preliminary
#molecule_type DNA
#residues 1-355 #label BAE
#cross-references GB:U18559; NID:g829370; PID:g829374
GENETICS lpfD
#gene #length 355 #molecular-weight 37714 #checksum 5448
SUMMARY

Query Match 70.6%; Score 48; DB 2; Length 355;
Best Local Similarity 63.6%; Pred. No. 3.61e+00;
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 51 TSKNTTGATFV 61
|||||:|
Qy 82 TSNDTTSAAFV 92

RESULT 13
ENTRY #A57172 #type complete
TITLE probable hormone receptor EMR1 precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change
24-Sep-1998
#accession A57172
REFERENCE A57172

#authors Baud, V.; Chissoe, S.L.; Viegas-Pequignot, E.; Diriong, S.;
N'Guyen, V.C.; Roe, B.A.; Lipinski, M.
#journal Genomics (1995) 26:334-344
#title EMR1, an unusual member in the family of hormone receptors
with seven transmembrane segments.
#accession A57172
#status preliminary
#molecule_type mRNA
#residues 1-886 #label BAU
#cross-references GB:X81479; NID:g784993; PID:g784994
GENETICS
#gene GDB:EMR1
#cross-references GDB:378349; OMIM:600493
#map_position 19p13.3-19p13.3
KEYWORDS transmembrane protein
SUMMARY #length 886 #molecular-weight 97679 #checksum 2055

Query Match 70.6%; Score 48; DB 2; Length 886;
Best Local Similarity 61.5%; Pred. No. 3.61e+00;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 470 SESTETGVAFVS 482
|||||:|
Qy 81 STSNDTTSAAFVS 93

RESULT 14
ENTRY #A34755 #type complete
TITLE nitrogen regulatory protein nit-2 - Neurospora crassa
ORGANISM #formal_name Neurospora crassa
DATE 13-Jul-1990 #sequence_revision 26-Jul-1996 #text_change
16-Feb-1997

ACCESSIONS A34755
REFERENCE A34755
#authors Fu, Y.H.; Marzluf, G.A.
#journal Mol. Cell. Biol. (1990) 10:1056-1065
#title nit-2, the major nitrogen regulatory gene of Neurospora
crassa, encodes a protein with a putative zinc finger
DNA-binding domain.
#cross-references MUID:90158568
#accession A34755
#molecule_type DNA; mRNA
#residues 1-1036 #label FUY
#cross-references GB:M33956
GENETICS
#introns 209/2; 335/3
CLASSIFICATION #superfamily nitrogen regulatory protein nit-2; GATA-type
zinc finger homology
KEYWORDS DNA binding; transcription regulation; zinc finger
FEATURE 740-793
#domain GATA-type zinc finger homology #label GZF
743-767
#region zinc finger GATA motif
SUMMARY #length 1036 #molecular-weight 109295 #checksum 235

Query Match 70.6%; Score 48; DB 1; Length 1036;
Best Local Similarity 53.8%; Pred. No. 3.61e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 87 SSSNNNTSNGFVA 99
|||||:|
Qy 81 STSNDTTSAAFVS 93

RESULT 15
ENTRY #S50536 #type complete
TITLE hypothetical protein YER033c - yeast (Saccharomyces
cerevisiae)
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 28-May-1993 #sequence_revision 24-Feb-1995 #text_change
21-Nov-1997
#accession S50536
REFERENCE S50428
#authors Dietrich, F.S.

```

```
REFERENCE      A41347
#authors      Ruth, P.; Rohrkasten, A.; Biel, M.; Bosse, E.; Regulla, S.;
#journal      Meyer, H.E.; Flockerzi, V.; Hofmann, F.
#title        Science (1989) 245:1115-1118
#keywords      Primary structure of the beta subunit of the DHP-sensitive
               calcium channel from skeletal muscle.
#cross-references MUID:89368946
#accession     A41347
#status        preliminary
#molecule_type mRNA
#residues      1-524 #label RUT
#cross-references GB:M25817
#keywords      skeletal muscle
#summary       #length 524 #molecular-weight 57868 #checksum 8438

Query Match      79.4%; Score 54; DB 2; Length 524;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches          9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNSFV 55
QY 81 STSNDTTSAAVF 92
    |||:||||:|

RESULT 7
ENTRY  myasthenic syndrome antigen B - human
TITLE  myasthenic syndrome antigen B - human
ALTERNATE_NAMES  MySB
ORGANISM  #formal_name Homo sapiens #common_name man
DATE 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change
20-Mar-1998
ACCESSIONS  A48895
REFERENCE  Rosenfeld, M.R.; Wong, E.; Dalmau, J.; Manley, G.; Posner,
#authors  J.B.; Sher, E.; Furneaux, H.M.
#journal  Ann. Neurol. (1993) 33:113-120
#title    Cloning and characterization of a Lambert-Eaton myasthenic
           syndrome antigen.
#cross-references MUID:93263585
#accession A48895
#status    preliminary
#molecule_type mRNA
#residues  1-567 #label ROS
#cross-references GB:S60415; NID:g300416; PID:g300417
#experimental_source fetal brain
#note      sequence extracted from NCBI backbone (NCBIN:132135,
           NCBIP:132136)
KEYWORDS  alternative splicing
SUMMARY   #length 567 #molecular-weight 62087 #checksum 9649

Query Match      79.4%; Score 54; DB 2; Length 567;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches          9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 58 STSDTTSNSFV 69
QY 81 STSNDTTSAAVF 92
    |||:||||:|

RESULT 8
ENTRY  voltage-dependent calcium channel beta subunit beta1B2
TITLE  isoform - human
ORGANISM  #formal_name Homo sapiens #common_name man
DATE 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change
20-Mar-1998
ACCESSIONS  C44461
REFERENCE  Powers, P.A.; Liu, S.; Hogan, K.; Gregg, R.G.
#authors  J. Biol. Chem. (1992) 267:22967-22972
#journal  Skeletal muscle and brain isoforms of a beta-subunit of human
#title     voltage-dependent calcium channels are encoded by a single
           gene.
```

```
#cross-references MUID:93054616
#accession     C44461
#status        preliminary; not compared with conceptual translation
#molecule_type nucleic acid
#residues      1-596 #label POW
#cross-references GB:M92303; NID:g291880; PID:g179806
#experimental_source hippocampus
#note          sequence extracted from NCBI backbone (NCBIP:118133)
SUMMARY       #length 596 #molecular-weight 65808 #checksum 4540

Query Match      79.4%; Score 54; DB 2; Length 596;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches          9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNSFV 55
QY 81 STSNDTTSAAVF 92
    |||:||||:|

RESULT 9
ENTRY  calcium channel protein beta chain - rat
TITLE  calcium channel protein beta chain - rat
ORGANISM  #formal_name Rattus norvegicus #common_name Norway rat
DATE 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
20-Mar-1998
ACCESSIONS  S18304
REFERENCE  Pragnell, M.; Sakamoto, J.; Jay, S.D.; Campbell, K.P.
#authors  FEBS Lett. (1991) 291:253-258
#journal  Cloning and tissue-specific expression of the brain calcium
#title     channel beta-subunit.
#cross-references MUID:92038046
#accession S18304
#molecule_type mRNA
#residues      1-597 #label PRA
#cross-references GB:X61394; NID:g55893; PID:g55894
SUMMARY       #length 597 #molecular-weight 65679 #checksum 8572

Query Match      79.4%; Score 54; DB 2; Length 597;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches          9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNSFV 55
QY 81 STSNDTTSAAVF 92
    |||:||||:|

RESULT 10
ENTRY  L-type voltage-gated calcium channel B subunit - human
TITLE  L-type voltage-gated calcium channel B subunit - human
ORGANISM  #formal_name Homo sapiens #common_name man
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
02-Jul-1996
ACCESSIONS  I52859
REFERENCE  Collin, T.; Wang, J.; Margeot, J.; Schwartz, A.
#authors  Circ. Res. (1993) 72:1337-1344
#journal  Molecular cloning of three isoforms of the L-type
#title     voltage-dependent calcium channel B subunit from normal
           human heart.
#cross-references MUID:93265672
#accession I52859
#status        preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues      1-597 #label RES
#cross-references GB:L06110; NID:g187014; PID:g187015
SUMMARY       #length 597 #molecular-weight 65578 #checksum 5631

Query Match      79.4%; Score 54; DB 2; Length 597;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches          9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNSFV 55
```

```

RESULT 2
ENTRY 165766 #type complete
TITLE L-type voltage-gated calcium channel B subunit - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
12-Aug-1996
ACCESSIONS 165766
REFERENCE Collin, T.; Wang, J.; Nargeot, J.; Schwartz, A.
#authors Circ. Res. (1993) 72:1337-1344
#journal Molecular cloning of three isoforms of the L-type
#title voltage-dependent calcium channel B subunit from normal
human heart.
#cross-references MUID:93265672
#accession 165766
#status preliminary; translated from GB/EMBL/DBBJ
#molecule_type mRNA
#residues 1-477 #label RES
#cross-references GB:L06111; NID:gl87016; PID:gl87017
SUMMARY #length 477 #molecular_weight 52977 #checksum 4957
Query Match 79.4%; Score 54; DB 2; Length 477;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 43 STSDTTNSFV 54
QY 81 STSNDTTSAAFV 92
|||||:|

RESULT 3
ENTRY JH0566 #type complete
TITLE calcium channel beta-2 chain - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
10-Sep-1997
ACCESSIONS JH0566
REFERENCE Williams, M.E.; Feldman, D.H.; McCue, A.F.; Brenner, R.;
#authors Vellicelebi, G.; Ellis, S.B.; Harpold, M.M.
#journal Neuron (1992) 8:71-84
#title Structure and functional expression of alpha1, alpha2, and
beta subunits of a novel human neuronal calcium channel
subtype.
#cross-references MUID:92110010
#accession JH0566
#molecule_type mRNA
#residues 1-478 #label WIL
#cross-references GB:M7560; NID:gl79743; PID:gl79744
#experimental_source hippocampus
COMMENT This protein is a subunit of the voltage dependent calcium channel.
KEYWORDS glycoprotein; phosphoprotein
FEATURE
32,167,209,348,374, #binding_site phosphate (Ser) (covalent) (by protein
450,464 kinase C) #status predicted\
64,201 #binding_site phosphate (Thr) (covalent) (by protein
kinase C) #status predicted\
189,425 #binding_site carbohydrate (Asn) (covalent) #status
predicted\
205 #binding_site phosphate (Thr) (covalent) (by
cAMP-dependent kinase) #status predicted
SUMMARY #length 478 #molecular_weight 52934 #checksum 6465
Query Match 79.4%; Score 54; DB 2; Length 478;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 44 STSDTTNSFV 55
QY 81 STSNDTTSAAFV 92
|||||:|

RESULT 4
ENTRY B44461 #type complete
TITLE voltage-dependent calcium channel beta subunit beta1B1
ORGANISM #formal_name Homo sapiens #common_name man
DATE 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change
20-Mar-1998
ACCESSIONS B44461
REFERENCE Powers, P.A.; Liu, S.; Hogan, K.; Gregg, R.G.
#authors J. Biol. Chem. (1992) 267:22967-22972
#journal Skeletal muscle and brain isoforms of a beta-subunit of human
#title voltage-dependent calcium channels are encoded by a single
gene.
#cross-references MUID:93054616
#accession B44461
#status preliminary
#molecule_type nucleic acid
#residues 1-478 #label POW
#cross-references GB:M92302; NID:gl79803; PID:gl79804
#experimental_source hippocampus
#note sequence inconsistent with the nucleotide translation
#note sequence extracted from NCBI backbone (NCBIP:118131)
SUMMARY #length 478 #molecular_weight 53176 #checksum 7630
Query Match 79.4%; Score 54; DB 2; Length 478;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 44 STSDTTNSFV 55
QY 81 STSNDTTSAAFV 92
|||||:|

RESULT 5
ENTRY I65767 #type complete
TITLE L-type voltage-gated calcium channel B subunit - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
12-Aug-1996
ACCESSIONS I65767
REFERENCE Collin, T.; Wang, J.; Nargeot, J.; Schwartz, A.
#authors Circ. Res. (1993) 72:1337-1344
#journal Molecular cloning of three isoforms of the L-type
#title voltage-dependent calcium channel B subunit from normal
human heart.
#cross-references MUID:93265672
#accession I65767
#status preliminary; translated from GB/EMBL/DBBJ
#molecule_type mRNA
#residues 1-522 #label RES
#cross-references GB:L06112; NID:gl87018; PID:gl87019
SUMMARY #length 522 #molecular_weight 57834 #checksum 7462
Query Match 79.4%; Score 54; DB 2; Length 522;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 43 STSDTTNSFV 54
QY 81 STSNDTTSAAFV 92
|||||:|

RESULT 6
ENTRY A41347 #type complete
TITLE calcium channel protein beta chain,
ORGANISM dihydropyridine-sensitive, skeletal muscle - rabbit
#formal_name Oryctolagus cuniculus #common_name domestic
#title rabbit
DATE 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change
07-Feb-1997
ACCESSIONS A41347

```

\*\*\*\*\*  
 W P S R H  
 (TW)  
 \*\*\*\*\*

Release 3.1A John F. Collins., Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Wed Sep 1 16:32:35 1999; MasPar time 4.57 Seconds  
 Tabular output not generated. 113.947 Million cell updates/sec

Title: >PCT-US99-13024-2  
 Description: (81-93) from PCTUS9913024.pep (12 of 12)  
 Perfect Score: 68  
 Sequence: 1 STSNDTTSAAAFVS 13

Scoring table: PAM 150  
 Gap 11  
 Searched: 122810 seqs, 40068593 residues  
 Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: pir60  
 1.pir1 2.pir2 3.pir3 4.pir4

Statistics: Mean 22.122; Variance 24.813; scale 0.892

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	54	79.4	475	2	A44461 voltage-dependent cal	1.44e-01
2	54	79.4	477	2	I65766 L-type voltage-gated	1.44e-01
3	54	79.4	478	2	JH0566 calcium channel beta-	1.44e-01
4	54	79.4	478	2	B44461 voltage-dependent cal	1.44e-01
5	54	78.4	522	2	I65767 L-type voltage-gated	1.44e-01
6	54	79.4	524	2	A41347 calcium channel prote	1.44e-01
7	54	79.4	567	2	A48895 myasthenic syndrome a	1.44e-01
8	54	79.4	596	2	A44461 voltage-dependent cal	1.44e-01
9	54	79.4	597	2	S18304 calcium channel prote	1.44e-01
10	54	79.4	597	2	I52859 L-type voltage-gated	1.44e-01
11	54	78.4	632	2	S21048 calcium channel prote	1.44e-01
12	48	70.6	355	2	D6271 long polar fibrillar o	3.61e+00
13	48	70.6	886	2	A57172 probable hormone rece	3.61e+00
14	48	70.6	1036	1	A34755 nitrogen regulatory p	3.61e+00
15	46	67.6	1076	2	S50536 hypothetcal protein	9.97e+00
16	45	66.2	345	2	A64370 adenylosuccinate synt	1.64e+01
17	45	66.2	416	2	S09761 hypothetcal protein	1.64e+01
18	45	66.2	583	2	S30959 probable membrane pro	1.64e+01
19	44	64.7	37	2	A21112 variant surface glyco	2.67e+01
20	44	64.7	236	2	I57681 potassium channel pro	2.67e+01
21	44	64.7	490	2	A35312 potassium channel pro	2.67e+01
22	44	64.7	588	2	A39624 probable helicase (EC	2.67e+01
23	44	64.7	1751	1	MMRHMH laminin alpha-2 chain	2.67e+01

## ALIGNMENTS

RESULT 1

ENTRY: A44461 #type complete  
 TITLE: voltage-dependent calcium channel beta subunit betalm isoform  
 ORGANISM: Homo sapiens #common\_name man  
 DATE: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change  
 10-Sep-1997  
 A44461; S31919  
 A44461  
 Powers, P.A.; Liu, S.; Hogan, K.; Gregg, R.G.  
 J. Biol. Chem. (1992) 267:22967-22972  
 Skeletal muscle and brain isoforms of a beta-subunit of human  
 voltage-dependent calcium channels are encoded by a single  
 gene.

#cross-references M01D:93054616

#accession A44461

#status preliminary

#molecule\_type nucleic acid

#residues 1-475 #label POM

#experimental\_source skeletal muscle

#note sequence inconsistent with the nucleotide translation

#note sequence extracted from NCBI backbone (NCBIP:118128)

REFERENCE S31919

Iles, D.E.; Segers, B.; Sengers, R.C.; Monsieurs, K.;

Heytens, L.; Halsall, J.; Hopkins, P.; Ellis, R.;

Hall-Curran, J.; Stewart, A.; Wieringa, B.

submitted to the EMBL Data Library, February 1993

Genetic mapping of the beta and gamma subunits of the L-type

voltage-dependent calcium channel on chromosome 17q and

exclusion as candidate genes for malignant hyperthermia

susceptibility.

accession S31919 preliminary

#status preliminary

#molecule\_type DNA

#residues 146-183 #label ILE

#cross-references EMBL:221725; NID:938562; PID:938563

SUMMARY #length 475 #molecular-weight 53079 #checksum 7282

Query Match 79.4%; Score 54; DB 2; Length 475;

Best Local Similarity 75.0%; Pred. No. 1.44e-01;

Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDDTTSNPFV 55

Qy 81 STSNDTTSAAFV 92

OS Homo sapiens.  
PN US5792846-A.  
PD 11-AUG-1998.  
PF 31-MAY-1995. 455543.  
PR 04-APR-1994; US-223305.  
PR 04-APR-1988; US-176899.  
PR 04-APR-1989; US-603751.  
PR 04-APR-1989; WO-001408.  
PR 20-FEB-1990; US-482384.  
PR 30-NOV-1990; US-620250.  
PR 15-AUG-1991; US-745206.  
PR 31-MAY-1995; US-455543.  
PA (SIBI-) SIBIA NEUROSCIENCES INC.  
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,  
PI Williams ME;  
DR WPI: 98-456192/39.  
DR N-PSDB: V42688.  
PT DNA encoding human calcium channel alpha 1B subunit protein -  
PT useful for recombinant production of the channel for screening of  
PT its modulators, and diagnosis of Lambert Eaton Syndrome  
PS Claim 3: Columns 219-224; 166pp; English.  
CC The present sequence represents a splice variant of the beta subunit of  
CC a human calcium channel. Calcium channels are membrane-spanning,  
CC multi-subunit proteins that allow controlled entry of calcium ions into  
CC cells. This leads to depolarisation events required for muscle  
CC contraction. The recombinant subunit, when expressed with nucleic acids  
CC encoding the complete calcium channel, can be used in assays for the  
CC detection and characterisation of compounds that modulate the channel.  
CC The DNA encoding the subunits can be alternatively spliced when  
CC transcribed, giving more than one form of the protein from the same  
CC transcript, each having slightly different properties. In addition, the  
CC reactivity of the alpha 1 subunit with IgG molecules from the serum of  
CC an individual with Lambert Eaton Syndrome (LES) can be used as a  
CC diagnostic for the disease.  
SQ Sequence 598 AA;

Query Match 79.4%; Score 54; DB 33; Length 598;  
Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnsv 55  
|||:||||:|  
Qy 81 STSNDTTSAAVF 92

RESULT 15  
ID R72612 standard; Protein: 598 AA.  
AC R72612;  
DT 01-DEC-1995 (first entry)  
DE Human calcium channel subunit beta 1-3.  
KW Calcium channel subunit; antagonist; agonist; diagnosis;  
KW Lambert Eaton Syndrome.  
OS Homo sapiens.  
PN WO9504822-A.  
PD 16-FEB-1995.  
PF 11-AUG-1994; U09230.  
PR 11-AUG-1993; US-105536.  
PR 05-NOV-1993; US-149097.  
PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.  
PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;  
DR WPI: 95-090900/12.  
DR N-PSDB: Q87839.  
PT DNA encoding human calcium channel sub-unit(s) - used for  
PT developing prods. for studying calcium channels, e.g. for  
PT obtaining agonists and antagonists  
PS Disclosure: Page 163-165; 283pp; English.  
CC To isolate DNA encoding the beta 1 subunit, a human hippocampus  
CC cDNA library was screened by hybridisation to a DNA fragment  
CC encoding a rabbit skeletal muscle calcium channel beta subunit.  
CC A hybridising clone was selected and was in turn used to isolate  
CC overlapping clones until the overlapping clones encompassing DNA  
CC encoding the entire human calcium channel beta 2 subunit were  
CC isolated and sequenced. Five alternatively spliced forms of the

CC beta 1 subunit have been identified. These forms are designated  
CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in the  
CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in  
CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293  
CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3  
CC subunits have been constructed. The subunits beta 1-1, beta 1-2,  
CC beta 1-4 and beta 1-5 have been identified by nucleic acid  
CC amplification, analysis as alternatively spliced forms of the  
CC beta subunit. Sequences of the beta 1 splice variants are set  
CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and  
CC R72604-R72706.  
SQ Sequence 598 AA;

Query Match 79.4%; Score 54; DB 14; Length 598;  
Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnsv 55  
|||:||||:|  
Qy 81 STSNDTTSAAVF 92

Search completed: Wed Sep 1 16:33:43 1999  
Job time : 36 secs.



```

KW Beta-1 subunit; human; calcium channel; assay; detection;
KW characterisation; Lambert Eaton Syndrome; LES; diagnosis.
OS Homo sapiens.
PN US5792846-A.
PD 11-AUG-1998.
PF 31-MAY-1995; 455543.
PR 04-APR-1994; US-223305.
PR 04-APR-1988; US-176899.
PR 04-APR-1989; US-603751.
PR 04-APR-1989; WO-U01408.
PR 20-FEB-1990; US-482384.
PR 30-NOV-1990; US-620250.
PR 15-AUG-1991; US-745206.
PR 31-MAY-1995; US-455543.
PA (SIBI-) SIBIA NEUROSCIENCES INC.
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
PI Williams ME;
PI WPI; 98-456192/39.
DR N-PSDB; V42705.
DR DNA encoding human calcium channel alpha 1B sub:unit protein -
PT useful for recombinant production of the channel for screening of
PT its modulators, and diagnosis of Lambert Eaton Syndrome
PS Claim 3; Columns 223-226; 166pp; English.
CC The present sequence represents the beta-1 subunit of a human calcium
CC channel. Calcium channels are membrane-spanning, multi-subunit proteins
CC that allow controlled entry of calcium ions into cells. This leads
CC to depolarisation events required for muscle contraction. The recombinant
CC subunit, when expressed with nucleic acids encoding the complete calcium
CC channel, can be used in assays for the detection and characterisation of
CC compounds that modulate the channel. The DNA encoding the subunits can
CC be alternatively spliced when transcribed, giving more than one form of
CC the protein from the same transcript, each having slightly different
CC properties. In addition, the reactivity of the alpha 1 subunit with IgG
CC molecules from the serum of an individual with Lambert Eaton Syndrome
CC (LES) can be used as a diagnostic for the disease.
SQ Sequence 523 AA;

Query Match 79.4%; Score 54; DB 33; Length 523;
Best Local Similarity 75.0%; Pred. No. 1.78e+01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdttsnsfv 55
QY 81 STSNDTTSAAVF 92

RESULT 12
ID W63147 standard; Protein; 530 AA.
AC W63147;
DT 12-OCT-1998 (first entry)
DE Human calcium channel beta subunit.
KW Beta subunit; human; calcium channel; assay; detection;
KW characterisation; Lambert Eaton Syndrome; LES; diagnosis.
OS Homo sapiens.
PN US5792846-A.
PD 11-AUG-1998.
PF 31-MAY-1995; 455543.
PR 04-APR-1994; US-223305.
PR 04-APR-1988; US-176899.
PR 04-APR-1989; US-603751.
PR 04-APR-1989; WO-U01408.
PR 20-FEB-1990; US-482384.
PR 30-NOV-1990; US-620250.
PR 15-AUG-1991; US-745206.
PR 31-MAY-1995; US-455543.
PA (SIBI-) SIBIA NEUROSCIENCES INC.
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
PI Williams ME;
PI WPI; 98-456192/39.
DR N-PSDB; V42693.
DR DNA encoding human calcium channel alpha 1B sub:unit protein -
PT useful for recombinant production of the channel for screening of
PT its modulators, and diagnosis of Lambert Eaton Syndrome

KW Beta-1 subunit; human; calcium channel; assay; detection;
KW characterisation; Lambert Eaton Syndrome; LES; diagnosis.
OS Homo sapiens.
PN US5792846-A.
PD 11-AUG-1998.
PF 31-MAY-1995; 455543.
PR 04-APR-1994; US-223305.
PR 04-APR-1988; US-176899.
PR 04-APR-1989; US-603751.
PR 04-APR-1989; WO-U01408.
PR 20-FEB-1990; US-482384.
PR 30-NOV-1990; US-620250.
PR 15-AUG-1991; US-745206.
PR 31-MAY-1995; US-455543.
PA (SIBI-) SIBIA NEUROSCIENCES INC.
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
PI Williams ME;
PI WPI; 98-456192/39.
DR N-PSDB; V42705.
DR DNA encoding human calcium channel alpha 1B sub:unit protein -
PT useful for recombinant production of the channel for screening of
PT its modulators, and diagnosis of Lambert Eaton Syndrome

PS Disclosure; Columns 125-130; 166pp; English.
CC The present sequence represents the beta subunit of a human calcium
CC channel. Calcium channels are membrane-spanning, multi-subunit
CC proteins that allow controlled entry of calcium ions into cells.
CC This leads to depolarisation events required for muscle contraction.
CC The recombinant subunit, when expressed with nucleic acids encoding
CC the complete calcium channel, can be used in assays for the detection
CC and characterisation of compounds that modulate the channel. The
CC DNA encoding the subunits can be alternatively spliced when
CC transcribed, giving more than one form of the protein from the same
CC transcript, each having slightly different properties. In addition, the
CC reactivity of the alpha 1 subunit with IgG molecules from the serum of
CC an individual with Lambert Eaton Syndrome (LES) can be used as a
CC diagnostic for the disease.
SQ Sequence 530 AA;

Query Match 79.4%; Score 54; DB 33; Length 530;
Best Local Similarity 75.0%; Pred. No. 1.78e+01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdttsnsfv 55
QY 81 STSNDTTSAAVF 92

RESULT 13
ID R39565 standard; Protein; 571 AA.
AC R39565;
DT 09-FEB-1994 (first entry)
DE Human neuronal VDCC beta-subunit encoded by clone HBB3.
KW Voltage-dependent calcium channel; VDCC; beta-subunit;
KW calcium flux; ss.
OS Homo sapiens.
PN DE4222126-A.
PD 19-AUG-1993.
PF 06-JUL-1992; 222126.
PR 17-FEB-1992; DE-204716.
PR 06-JUL-1992; DE-222126.
PA (FARB ) BAYER AG.
PI Spreyer P, Unterbeck A;
PI WPI; 93-265734/34.
DR N-PSDB; Q46078.
DR Human neuronal beta-unit cDNA of voltage dependent calcium
PT channels - useful in calcium-flux studies and screening systems
PT for agonists and antagonists of calcium channels
PS Claim 2; Page 11-13; 13pp; German.
CC A first oligonucleotide probe (Q46075) complementary to nucleotides
CC 361-400 of the VDCC beta-subunit from rabbit skeletal muscle was
CC used to screen a human hippocampus cDNA library. A 1.9kb cDNA
CC fragment was isolated for further screening to isolate human VDCC
CC beta-subunit coding sequences. Clone HBB3 was sequenced and
CC nucleotides 1-1288 were found to have 92% homology to the rabbit
CC beta-subunit; from position 1289 there is no detectable homology.
CC The amino acid sequence R39565 was deduced from the open reading
CC frame. See also Q46076 and Q46077.
SQ Sequence 571 AA;

Query Match 79.4%; Score 54; DB 8; Length 571;
Best Local Similarity 75.0%; Pred. No. 1.78e+01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdttsnsfv 55
QY 81 STSNDTTSAAVF 92

RESULT 14
ID W63144 standard; Protein; 598 AA.
AC W63144;
DT 12-OCT-1998 (first entry)
DE Human calcium channel beta subunit splice variant beta3.
KW Beta subunit; human; calcium channel; assay; detection;
KW characterisation; Lambert Eaton Syndrome; LES; diagnosis.

```

RESULT 8  
 ID R72611 standard; Protein; 478 AA.  
 AC R72611:  
 DT 01-DEC-1995 (first entry)  
 DE Human calcium channel subunit beta 1.  
 KW Calcium channel subunit; antagonist; agonist; diagnosis;  
 KW Lambert Eaton Syndrome.  
 OS Homo sapiens.  
 PN W09504822-A.  
 PD 16-FEB-1995.  
 PF 11-AUG-1994; U09230.  
 PR 11-AUG-1993; US-105536.  
 PR 05-NOV-1993; US-149097.  
 PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.  
 PI Ellis SB, Gillespie A, Harpold MM, Mccue AF, Williams ME;  
 DR WPI: 95-090900/12.  
 DR N-PSDB: Q87838.  
 PT DNA encoding human calcium channel sub-unit(s) - used for  
 PT developing prods. for studying calcium channels, e.g. for  
 PT obtaining agonists and antagonists  
 PS Disclosure: Page 160-162; 285pp; English.  
 CC To isolate DNA encoding the Beta 1 subunit, a human hippocampus  
 CC cDNA library was screened by hybridisation to a DNA fragment  
 CC encoding a rabbit skeletal muscle calcium channel beta subunit.  
 CC A hybridising clone was selected and was in turn used to isolate  
 CC overlapping clones until the overlapping clones encompassing DNA  
 CC encoding the entire human calcium channel beta 2 subunit were  
 CC isolated and sequenced. Five alternatively spliced forms of the  
 CC beta 1 subunit have been identified. These forms are designated  
 CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in the  
 CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in  
 CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293  
 CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3  
 CC subunits have been constructed. The subunits beta 1-1, beta 1-2,  
 CC beta 1-4 and beta 1-5 have been identified by nucleic acid  
 CC amplification analysis as alternatively spliced forms of the  
 CC beta subunit. Sequences of the beta 1 splice variants are set  
 CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and  
 CC R72604-R72706.  
 SQ Sequence 478 AA;  
 Query Match 79.4%; Score 54; DB 14; Length 478;  
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 44 stsdtdttsnsvf 55  
 |||:||||:|  
 QY 81 STSNDTTSAAVF 92  
 RESULT 9  
 ID R33551 standard; Protein; 478 AA.  
 AC R33551:  
 DT 30-JUN-1993 (first entry)  
 DE Sequence of splice variant beta 1-2 of beta human calcium  
 DE channel subunit.  
 KW Human calcium channel subunit; diagnosis; agonist; antagonist;  
 KW Lambert Eaton syndrome.  
 OS Homo sapiens.  
 PN W09304083-A.  
 PD 04-MAR-1993.  
 PF 14-AUG-1992; U06903.  
 PR 15-AUG-1991; US-745206.  
 PR 10-APR-1992; US-868354.  
 PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.  
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, Mccue AF,  
 PI Williams ME;  
 DR WPI: 93-093836/11.  
 DR N-PSDB: Q37819.  
 PT DNA encoding specific human calcium channel sub-units - used for  
 PT identifying calcium channel agonists and antagonists and  
 PT diagnosing Lambert Eaton syndrome

PS Disclosure: Page 129-131; 150pp; English.  
 CC Five alternatively spliced forms of the human calcium channel  
 CC beta 1 subunit have been identified and DNA encoding a number 1-1,  
 CC of forms have been isolated. These forms are designated beta 1-1,  
 CC expressed in skeletal muscle, beta 1-2, expressed in the CNS, beta  
 CC 1-3, also expressed in the CNS, beta 1-4, expressed in aorta tissue  
 CC and HEK 293 cells, and beta 1-5, expressed in HEK 293 cells.  
 SQ Sequence 478 AA;  
 Query Match 79.4%; Score 54; DB 6; Length 478;  
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 44 stsdtdttsnsvf 55  
 |||:||||:|  
 QY 81 STSNDTTSAAVF 92  
 RESULT 10  
 ID R72604 standard; Protein; 479 AA.  
 AC R72604:  
 DT 01-DEC-1995 (first entry)  
 DE Human calcium channel subunit beta 1-1.  
 KW Calcium channel subunit; antagonist; agonist; diagnosis;  
 KW Lambert Eaton Syndrome.  
 OS Homo sapiens.  
 PN W09504822-A.  
 PD 16-FEB-1995.  
 PF 11-AUG-1994; U09230.  
 PR 11-AUG-1993; US-105536.  
 PR 05-NOV-1993; US-149097.  
 PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.  
 PI Ellis SB, Gillespie A, Harpold MM, Mccue AF, Williams ME;  
 DR WPI: 95-090900/12.  
 DR N-PSDB: Q87831.  
 PT DNA encoding human calcium channel sub-unit(s) - used for  
 PT developing prods. for studying calcium channels, e.g. for  
 PT obtaining agonists and antagonists  
 PS Disclosure: Page 253-256; 285pp; English.  
 CC To isolate DNA encoding the Beta 1 subunit, a human hippocampus  
 CC cDNA library was screened by hybridisation to a DNA fragment  
 CC encoding a rabbit skeletal muscle calcium channel beta subunit.  
 CC A hybridising clone was selected and was in turn used to isolate  
 CC overlapping clones until the overlapping clones encompassing DNA  
 CC encoding the entire human calcium channel beta 2 subunit were  
 CC isolated and sequenced. Five alternatively spliced forms of the  
 CC beta 1 subunit have been identified. These forms are designated  
 CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in the  
 CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in  
 CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293  
 CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3  
 CC subunits have been constructed. The subunits beta 1-1, beta 1-2,  
 CC beta 1-4 and beta 1-5 have been identified by nucleic acid  
 CC amplification analysis as alternatively spliced forms of the  
 CC beta subunit. Sequences of the beta 1 splice variants are set  
 CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and  
 CC R72604-R72706.  
 SQ Sequence 479 AA;  
 Query Match 79.4%; Score 54; DB 14; Length 479;  
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 44 stsdtdttsnsvf 55  
 |||:||||:|  
 QY 81 STSNDTTSAAVF 92  
 RESULT 11  
 ID W63156 standard; Protein; 523 AA.  
 AC W63156;  
 DT 12-OCT-1998 (first entry)  
 DE Human calcium channel beta-1 subunit.

CC be alternatively spliced when transcribed, giving more than one form of  
 CC the protein from the same transcript, each having slightly different  
 CC properties. In addition, the reactivity of the alpha 1 subunit with IgG  
 CC molecules from the serum of an individual with Lambert Eaton Syndrome  
 CC (LES) can be used as a diagnostic for the disease.  
 SQ Sequence 219 AA;

Query Match 79.4%; Score 54; DB 33; Length 219;

Best Local Similarity 75.0%; Pred. No. 1.78e+01;

Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnsv 55

||||| :||

QY 81 STSNDTTSAAVF 92

RESULT 5

ID R39697 standard; Protein; 240 AA.

AC R39697;

DT 19-JAN-1994 (first entry)

DE Myasthenic antigenic polypeptide.

KW MAP: antibody detection; LEMS; Lambert-Eaton myasthenic syndrome;

OS paraneoplastic sensory neuropathy; p mys B; mysB; ss.

PN Homo sapiens.

PD WO9314098-A.

PP 22-JUL-1993.

PR 11-JAN-1993; U00227.

PR 10-JAN-1992; US-820312.

PA (SLOK ) SLOAN KETTERING INST CANCER.

PI Furneaux HM, Posner JB;

DR WPI; 93-243126/30.

DR N-PSDB; Q46673.

PT New purified myasthenic antigenic polypeptide and its corresp.

PT antibody - useful for diagnosing and treating proliferation of

PT neoplastic cells in patient with Lambert-Eaton myasthenic syndrome

PS Claim 5; Page 26-27; 48pp; English.

CC The sequence is that of myasthenic antigenic polypeptide (MAP)

CC encoded by the cDNA clone p mysB. MAP can be used to detect

CC antibodies associated with paraneoplastic sensory neuropathy

CC such as Lambert-Eaton myasthenic syndrome (LEMS). These antibodies

CC are used to determine if a patient with neurological symptoms has a

CC tumour expressing MAP, to inhibit proliferation of neoplastic cells

CC in patients with LEMS and for imaging neoplastic cells in LEMS

SQ patients. 240 AA;

Query Match 79.4%; Score 54; DB 8; Length 240;

Best Local Similarity 75.0%; Pred. No. 1.78e+01;

Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 17 stsdtttsnsv 28

||||| :||

QY 81 STSNDTTSAAVF 92

RESULT 6

ID R39563 standard; Protein; 478 AA.

AC R39563;

DT 09-FEB-1994 (first entry)

DE Human neuronal VDCC beta-subunit encoded by clone HB1.

KW Voltage-dependent calcium channel; VDCC; beta-subunit;

KW calcium flux; ss.

OS Homo sapiens.

PN DE422126-A.

PP 19-AUG-1993.

PF 06-JUL-1992; 222126.

PR 17-FEB-1992; DE-204716.

PR 06-JUL-1992; DE-222126.

PA (FARB ) BAYER AG.

PI Spreyer P, Unterbeck A;

DR WPI; 93-265734/34.

DR N-PSDB; Q46076.

PT Human neuronal beta-unit cDNA of voltage dependent calcium

PT channels - useful in calcium-flux studies and screening systems  
 PT for agonists and antagonists of calcium channels  
 PS Claim 2; Page 5-7; 13pp; German  
 CC A first oligonucleotide probe (Q46075) complementary to nucleotides  
 CC 361-400 of the VDCC beta-subunit from rabbit skeletal muscle was  
 CC used to screen a human hippocampus cDNA library. A 1.9kb cDNA  
 CC fragment was isolated for further screening to isolate human VDCC  
 CC beta-subunit coding sequences. Clone HB1 was sequenced and found  
 CC to have 92% homology to the rabbit beta-subunit. The amino acid  
 CC sequence R39563 was deduced from the open reading frame.  
 CC See also Q46077-Q46078.  
 SQ Sequence 478 AA;

Query Match 79.4%; Score 54; DB 8; Length 478;

Best Local Similarity 75.0%; Pred. No. 1.78e+01;

Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnsv 55

||||| :||

QY 81 STSNDTTSAAVF 92

RESULT 7

ID W63143 standard; Protein; 478 AA.

AC W63143;

DT 12-OCT-1998 (first entry)

DE Human calcium channel beta subunit splice variant beta2.

KW Beta subunit; human; calcium channel; assay; detection;

KW characterisation; Lambert Eaton Syndrome; LES; diagnosis.

OS Homo sapiens.

PN US5792846-A.

PD 11-AUG-1998.

PP 31-MAY-1995; 455543.

PR 04-APR-1994; US-223305.

PR 04-APR-1988; US-176899.

PR 04-APR-1989; US-603751.

PR 04-APR-1989; WO-U01408.

PR 20-FEB-1990; US-482384.

PR 30-NOV-1990; US-620250.

PR 15-AUG-1991; US-745206.

PR 31-MAY-1995; US-455543.

PA (SIBI-) SIBIA NEUROSCIENCES INC.

PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,

PI Williams ME;

DR WPI; 98-456192/39.

DR N-PSDB; V42687.

PT DNA encoding human calcium channel alpha 1B subunit protein -

PT useful for recombinant production of the channel for screening of

PT its modulators, and diagnosis of Lambert Eaton Syndrome

PS Claim 3; Columns 217-220; 166pp; English.

CC The present sequence is encoded by a splice variant of the beta subunit

CC of a human calcium channel. Calcium channels are membrane-spanning,

CC multi-subunit proteins that allow controlled entry of calcium ions into

CC cells. This leads to depolarisation events required for muscle

CC contraction. The recombinant subunit, when expressed with nucleic acids

CC encoding the complete calcium channel, can be used in assays for the

CC detection and characterisation of compounds that modulate the channel.

CC The DNA encoding the subunits can be alternatively spliced when

CC transcribed, giving more than one form of the protein from the same

CC transcript, each having slightly different properties. In addition, the

CC reactivity of the alpha 1 subunit with IgG molecules from the serum of

CC an individual with Lambert Eaton Syndrome (LES) can be used as a

CC diagnostic for the disease.  
 SQ Sequence 478 AA;

Query Match 79.4%; Score 54; DB 33; Length 478;

Best Local Similarity 75.0%; Pred. No. 1.78e+01;

Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnsv 55

||||| :||

QY 81 STSNDTTSAAVF 92

```

SQ Sequence 216 AA;
Query Match 79.4%; Score 54; DB 33; Length 216;
Best Local Similarity 75.0%; Pred. No. 1.78e+01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfsv 55
   |||:||||:|
QY 81 STSNDTTSAAVF 92

RESULT 2
ID R72605 standard; Protein; 216 AA.
AC R72605;
DT 01-DEC-1995 (first entry)
DE Human calcium channel subunit beta 1-4.
KW Calcium channel subunit; antagonist; agonist; diagnosis;
KW Lambert Eaton Syndrome.
OS Homo sapiens.
PN W09504822-A.
PD 16-FEB-1995.
PF 11-AUG-1994; U09230.
PF 11-AUG-1993; US-105536.
PR 05-NOV-1993; US-149097.
PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.
PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;
DR WPI: 95-090900/12.
DR N-PSDB; Q87832.

PT DNA encoding human calcium channel sub-unit(s) - used for
PT developing prods. for studying calcium channels, e.g. for
PT obtaining agonists and antagonists
PS Disclosure; Page 256-257; 285pp; English.
CC cDNA library was screened by hybridisation to a DNA fragment
CC encoding a rabbit skeletal muscle calcium channel beta subunit.
CC A hybridising clone was selected and was in turn used to isolate
CC overlapping clones until the overlapping clones encompassing DNA
CC encoding the entire human calcium channel beta 2 subunit were
CC isolated and sequenced. Five alternatively spliced forms of the
CC beta 1 subunit have been identified. These forms are designated
CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in
CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in
CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293
CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3
CC subunits have been constructed. The subunits beta 1-1, beta 1-2,
CC beta 1-4 and beta 1-5 have been identified by nucleic acid
CC amplification analysis as alternatively spliced forms of the
CC beta subunit. Sequences of the beta 1 splice variants are set
CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and
CC R72604-R72706.
SQ Sequence 219 AA;

Query Match 79.4%; Score 54; DB 14; Length 219;
Best Local Similarity 75.0%; Pred. No. 1.78e+01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfsv 55
   |||:||||:|
QY 81 STSNDTTSAAVF 92

RESULT 4
ID W63158 standard; Protein; 219 AA.
AC W63158;
DT 12-OCT-1998 (first entry)
DE Human calcium channel beta-5 subunit.
KW Beta-5 subunit; human; calcium channel; assay; detection;
KW characterisation; Lambert Eaton Syndrome; LES; diagnosis; ds.
OS Homo sapiens.
PN U55792846-A.
PD 11-AUG-1998.
PF 31-MAY-1995; 455543.
PF 04-APR-1994; US-223305.
PR 04-APR-1988; US-176899.
PR 04-APR-1989; US-603751.
PR 04-APR-1989; WO-001408.
PR 20-FEB-1990; US-482384.
PR 30-NOV-1990; US-620250.
PR 15-AUG-1991; US-745206.
PR 31-MAY-1995; US-455543.
PA (SIBI-) SIBIA NEUROSCIENCES INC.
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
PI Williams ME;
DR WPI: 98-456192/39.
DR N-PSDB; V42707.

PT DNA encoding human calcium channel alpha 1B subunit protein -
PT useful for recombinant production of the channel for screening of
PT its modulators, and diagnosis of Lambert Eaton Syndrome
PS Disclosure; Columns 227-230; 166pp; English.
CC The present sequence represents the beta-5 subunit of a human calcium
CC channel. Calcium channels are membrane-spanning, multi-subunit proteins
CC that allow controlled entry of calcium ions into cells. This leads
CC to depolarisation events required for muscle contraction. The recombinant
CC subunit, when expressed with nucleic acids encoding the complete calcium
CC channel, can be used in assays for the detection and characterisation of
CC compounds that modulate the channel. The DNA encoding the subunits can

```

\*\*\*\*\*  
 W P S R E L H  
 \*\*\*\*\*  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:33:07 1999; MasPar time 4.43 Seconds

Tabular output not generated. 62.340 Million cell updates/sec

Title: >PCT-US99-13024-2

Description: (81-93) from PCTUS9913024.pep (12 of 12)

Sequence: 1 STSNDTTSAAAFVS 13

Scoring table: PAM 150  
 Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21 22:part22 23:part23  
 24:part24 25:part25 26:part26 27:part27 28:part28  
 29:part29 30:part30 31:part31 32:part32 33:part33  
 34:part34 35:part35 36:part36 37:part37 38:part38  
 39:part39

Statistics: Mean 15.573; Variance 46.313; scale 0.336

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	54	79.4	216	33	Human calcium channel	1.78e+01
2	54	79.4	216	14	Human calcium channel	1.78e+01
3	54	79.4	219	14	Human calcium channel	1.78e+01
4	54	79.4	219	33	Human calcium channel	1.78e+01
5	54	79.4	240	8	Myasthenic antigenic	1.78e+01
6	54	79.4	478	8	Human neuronal VDCC b	1.78e+01
7	54	79.4	478	33	Human calcium channel	1.78e+01
8	54	79.4	478	14	Human calcium channel	1.78e+01
9	54	79.4	478	6	Sequence of splice va	1.78e+01
10	54	79.4	479	14	Human calcium channel	1.78e+01
11	54	79.4	523	33	Human calcium channel	1.78e+01
12	54	79.4	530	33	Human calcium channel	1.78e+01
13	54	79.4	571	8	Human neuronal VDCC b	1.78e+01
14	54	79.4	578	33	Human calcium channel	1.78e+01
15	54	79.4	581	14	Human calcium channel	1.78e+01
16	54	79.4	581	6	Sequence of splice va	1.78e+01

17	54	79.4	660	14	R72613	Human neuronal calcium	1.78e+01
18	44	64.7	636	36	W79589	Human Kv potassium ch	2.29e+02
19	44	64.7	636	36	W79591	Human Kv potassium ch	2.29e+02
20	44	64.7	655	36	W79590	Human Kv potassium ch	2.29e+02
21	44	64.7	1130	15	R71729	Merosin major subunit	2.29e+02
22	44	64.7	1130	3	R13436	Merosin M polypeptide	2.29e+02
23	44	64.7	3110	15	R71730	Merosin major subunit	2.29e+02
24	43	63.2	757	20	W03179	Bovine poly-immunoglob	2.93e+02
25	42	61.8	127	28	W31716	Gamma-1 heavy chain a	3.74e+02
26	42	61.8	653	21	W14264	2. japonica phosphoen	3.74e+02
27	42	61.8	953	19	R97284	Human 26S proteasome	3.74e+02
28	42	61.8	2783	5	R23963	AFP-1 (Ala 2460 Val)	3.74e+02
29	42	61.8	2783	5	R23962	AFP-1.	3.74e+02
30	41	60.3	98	13	R72068	DP10 VH region.	4.76e+02
31	41	60.3	120	25	W27550	Human Ab heavy chain	4.76e+02
32	41	60.3	120	10	R54795	SPA-reactive IgM heav	4.76e+02
33	41	60.3	123	24	W19881	CEA-specific antibody	4.76e+02
34	41	60.3	123	24	W19889	CEA-specific antibody	4.76e+02
35	41	60.3	123	24	W19888	CEA-specific antibody	4.76e+02
36	41	60.3	123	24	W19887	CEA-specific antibody	4.76e+02
37	41	60.3	128	28	W31715	Gamma-1 heavy chain a	4.76e+02
38	41	60.3	268	27	W27136	Achromobacter lyticus	4.76e+02
39	41	60.3	268	27	W31403	Achromobacter lyticus	4.76e+02
40	41	60.3	268	26	W27135	Achromobacter lyticus	4.76e+02
41	41	60.3	397	19	W04270	B.t. alkaline proteas	4.76e+02
42	41	60.3	481	5	R24442	Sequence of antibody	4.76e+02
43	41	60.3	512	39	W87797	Protease encoded by c	4.76e+02
44	41	60.3	512	14	R80505	S. lividans protease	4.76e+02
45	41	60.3	1170	12	R63231	Crystal protein CryET	4.76e+02

## ALIGNMENTS

RESULT 1  
 ID W63157 standard; Protein; 216 AA.

DE 12-OCT-1998 (first entry)

DT Human calcium channel beta-4 subunit.

KW Beta-4 subunit; human; calcium channel; assay; detection;

OS Homo sapiens.

PN US5792846-A.

PD 11-AUG-1998.

PF 31-MAY-1995; 455543.

PR 04-APR-1994; US-223305.

PR 04-APR-1988; US-176899.

PR 04-APR-1989; US-603751.

PR 04-APR-1989; WO-001408.

PR 20-FEB-1990; US-482384.

PR 30-NOV-1990; US-620250.

PR 15-AUG-1991; US-745206.

PR 31-MAY-1995; US-455543.

PA (SIBI-) SIBIA NEUROSCIENCES INC.

PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,

PI Williams ME;

DR WPI: 98-456192/39.

DR N-PSDB: V43706.

PT DNA encoding human calcium channel alpha 1B subunit protein -

PT useful for recombinant production of the channel for screening of

PT its modulators, and diagnosis of Lambert Eaton Syndrome

PS Disclosure: Columns 225-228; 166pp; English.

CC The present sequence represents the beta-4 subunit of a human calcium

CC channel. Calcium channels are membrane-spanning, multi-subunit proteins

CC that allow controlled entry of calcium ions into cells. This leads

CC to depolarisation events required for muscle contraction. The recombinant

CC subunit, when expressed with nucleic acids encoding the complete calcium

CC channel, can be used in assays for the detection and characterisation of

CC compounds that modulate the channel. The DNA encoding the subunits can

CC be alternatively spliced when transcribed, giving more than one form of

CC the protein from the same transcript, each having slightly different

CC properties. In addition, the reactivity of the alpha 1 subunit with IgG

CC molecules from the serum of an individual with Lambert Eaton Syndrome

CC (LES) can be used as a diagnostic for the disease.

Query Match 58.1%; Score 50; DB 2; Length 371;  
Best Local Similarity 53.8%; Pred. No. 9.22e+00;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 164 DGTGNLIQSSY 176  
QY 66 DPSTGALVDSKY 78

RESULT 14  
ID O70461 PRELIMINARY; PRT; 492 AA.  
AC O70461;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE MONOCARBOXYLATE TRANSPORTER MCT3.  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-RETINA;  
RA PHILP N.J., YOON H.;  
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AF059258; G3170609;  
SQ SEQUENCE 492 AA; 51590 MW; 20F051C1 CRC32;

Query Match 58.1%; Score 50; DB 11; Length 492;  
Best Local Similarity 61.5%; Pred. No. 9.22e+00;  
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 373 PSAGRLVDALKNY 385  
QY 67 PSTGALVDS-KSY 78

RESULT 15  
ID O13151 PRELIMINARY; PRT; 542 AA.  
AC O13151;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE MONOCARBOXYLATE TRANSPORTER 3.  
GN MCT3.  
OS GALLUS GALLUS (CHICKEN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LEGHORN;  
RA PHILP N.J., YOON H.;  
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AF000240; G2198807;  
SQ SEQUENCE 542 AA; 58085 MW; 836DA7B1 CRC32;

Query Match 58.1%; Score 50; DB 13; Length 542;  
Best Local Similarity 61.5%; Pred. No. 9.22e+00;  
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 410 PSAGRLVDALKNY 422  
QY 67 PSTGALVDS-KSY 78

Search completed: Wed Sep 1 16:29:45 1999  
Job time : 21 secs.

```

QY 66 DPSTGALVDSKSYA 79
      || |||::| |::|
RESULT 10 PRELIMINARY; PRT; 142 AA.
ID Q87600;
AC Q87600;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS SIMIAN IMMUNODEFICIENCY VIRUS (SIV-AGM).
OC VIRUSES; RETROVIRUSES; RETROVIRIDAE; LENTIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P081;
RA BIBOLLET-RUCHE F., BENGUES C., GALAT-LUONG A., GALAT G., POURRUT X.,
RA VIDAL N., VEAS F., DURAND J.P., CUNY G.;
RL J. VIROL. 71:307-313(1997).
DR EMBL; U37209; G1353456; -.
DR EMBL; U37209; G1353456; -.
KW ENVELOPE PROTEIN.
DR PFAM; PF00517; GP41; 1.
FT NON_TER 1
FT NON_TER 142
FT NON_TER 142
SQ SEQUENCE 142 AA; 16852 MW; C58A9A91 CRC32;

Query Match 58.1%; Score 50; DB 14; Length 142;
Best Local Similarity 46.7%; Pred. No. 9.22e+00;
Matches 7; Conservative 5; Mismatches 2; Indels 1; Gaps 1;

Db 86 EPDSGGLSRDSRSY 100
      |::|::|::|
QY 66 DPSTGALV-DSKSYA 79

RESULT 11 PRELIMINARY; PRT; 156 AA.
ID Q35308;
AC Q35308;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE MONOCARBOXYLATE TRANSPORTER MCT3 (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H;
RA PHILP N.J., YOON H.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF019111; G2407664; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 156 AA; 16253 MW; DB394B9E CRC32;

Query Match 58.1%; Score 50; DB 11; Length 156;
Best Local Similarity 61.5%; Pred. No. 9.22e+00;
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 37 PSAGRLVDALKNY 49
      |||::|::|::|
QY 67 PSTGALVDS-KSY 78

RESULT 12 PRELIMINARY; PRT; 262 AA.
ID Q44188;
AC Q44188;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE W-AMINO-TRANSFERASE-LIKE PROTEIN.
GN OATA.
OS AGROBACTERIUM RADIOBACTER.

```

```

OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; AGROBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T1305 LAC9;
RX MEDLINE; 96425889.
RA TIBURTIUS A., DE LUCA N., HUSSAIN H., JOHNSTON A.W.B.;
RT "Expression of the exy gene, required for exopolysaccharide
RT synthesis in Agrobacterium, is activated by the regulatory ros
RT gene.";
RL MICROBIOLOGY 142:2621-2629(1996).
DR EMBL; X95394; E220383; -.
DR PFAM; PF00202; aminotran_3; 1.
KW TRANSFERASE.
SQ SEQUENCE 262 AA; 28124 MW; DF7A4568 CRC32;

Query Match 58.1%; Score 50; DB 2; Length 262;
Best Local Similarity 63.6%; Pred. No. 9.22e+00;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 110 STGVLVPPKGY 120
      |||::|::|
QY 68 STGALVDSKSY 78

RESULT 13 PRELIMINARY; PRT; 371 AA.
ID P72097;
AC P72097; P72099; P72100; P72101;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CMP-N-ACETYLNEURAMINATE-BETA-GALACTOSAMIDE-ALPHA-2,3-
DE SIALYLTRANSFERASE (EC 2.4.99.-)
DE (BETA-GALACTOSIDE ALPHA-2,3-SIALYLTRANSFERASE) (ALPHA 2,3-ST)
DE (LIPOLIGOSACCHARIDE SIALYLTRANSFERASE) (LST).
GN LST.
OS NEISSERIA MENINGITIDIS.
OC BACTERIA; PROTEOBACTERIA; BETA SUBDIVISION; NEISSERIACEAE; NEISSERIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / NRCC 4728, 406Y / NRCC 4030, AND M982B / NRCC 4725;
RA GILBERT M., WATSON D.C., CUNNINGHAM A.-M., JENNINGS M.P., YOUNG N.M.,
RA WAKARCHUK W.W.;
J. BIOL. CHEM. 271:28271-28276(1996).
CC -1- FUNCTION: TRANSFERS SIALIC ACID FROM THE SUBSTRATE CMP-SIALIC
CC ACID DONOR TO THE TERMINAL
CC BETA-D-GALACTOSYL-1,4-ACETYL-BETA-D-GLUCOSAMINE ON THE
CC LACTO-N-NEOTETRAOSE BRANCH OF THE LIPOLIGOSACCHARIDE.
CC -1- CATALYTIC ACTIVITY: CMP-N-ACETYLNEURAMINATE +
CC BETA-D-GALACTOSYL-1,4-ACETYL-BETA-D-GLUCOSAMINE -> CMP +
CC ALPHA-N-ACETYLNEURAMINYL-2,3-BETA-D-GALACTOSYL-1,4-N-ACETYL-BETA-
CC D-GLUCOSAMINE.
CC -1- PATHWAY: GLYCOSYLATION.
CC EMBL; U60660; G1546004; -.
DR EMBL; U60661; G1546007; -.
DR EMBL; U60662; G1546009; -.
DR EMBL; U60663; G1546011; -.
KW TRANSFERASE; GLYCOSYLTRANSFERASE.
FT VARIANT 2 2
FT VARIANT 29 29 G -> S (IN STRAIN M982B / NRCC 4725).
FT (IN STRAINS 406Y / NRCC 4030 AND M982B /
FT NRCC 4725).
FT E -> D
FT (IN STRAINS 406Y / NRCC 4030 AND M982B /
FT NRCC 4725).
FT N -> K
FT (IN STRAINS 406Y / NRCC 4030 AND M982B /
FT NRCC 4725).
FT L -> *; (IN STRAIN M982B / NRCC 4725;
FT LOSS OF ACTIVITY).
FT T -> A (IN STRAIN 406Y / NRCC 4030).
FT K -> N (IN STRAIN 406Y / NRCC 4030).
FT SEQUENCE 371 AA; 42611 MW; 73DB83A9 CRC32;

```

```

RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 98290545.
RA NAGASE T., ISHIKAWA K., MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N.,
RA OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RL code for large proteins in vitro.";
RL DNA RES. 5:31-39(1998).
DR EMBL; AB011135; D1026419; -.
SQ SEQUENCE 870 AA; 94665 MW; EB9CD3F4 CRC32;

Query Match 60.5%; Score 52; DB 4; Length 870;
Best Local Similarity 42.9%; Pred. No. 3.35e+00;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 798 EPAQDLSLVQSESYT 811
   |||:|||||
QY 66 DPSTGALVDSKSYA 79

RESULT 7
ID Q24281 PRELIMINARY; PRT; 540 AA.
AC Q24281;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOTOLYASE.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PHR.; TISSUE=OVARY;
RX MEDLINE; 96178677.
RA TODO T., RYO H., YAMAMOTO K., TOH H., INUI T., AYAKI H., NOMURA T.,
RA IKENAGA M.;
RT "Similarity among the Drosophila (6-4)photolyase, a human photolyase
RT homolog, and the DNA photolyase-blue-light photoreceptor family.";
RL SCIENCE 272:109-112(1996).
DR EMBL; D83701; D1012739; -.
KW LYASE.
SQ SEQUENCE 540 AA; 62548 MW; 62F139EE CRC32;

Query Match 59.3%; Score 51; DB 5; Length 540;
Best Local Similarity 35.7%; Pred. No. 5.58e+00;
Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 456 EPKASLVQORAYG 469
   .||:|||||:|
QY 66 DPSTGALVDSKSYA 79

RESULT 8
ID Q88274 PRELIMINARY; PRT; 1340 AA.
AC Q88274;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE POLYPROTEIN (FRAGMENT).
OS SWEET POTATO FEATHERY MOTTLE VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC POTYVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEVERE;
RX MEDLINE; 95390803.
RA MORI M., SAKAI J., KIMURA T., USUGI T., HAYASHI T., HANADA K.,
RA NISHIGUCHI M.;
RT "Nucleotide sequence analysis of two nuclear inclusion body and coat
RT protein genes of a sweet potato feathery mottle virus severe strain
RT (SPFMV-S) genomic RNA.";

```

```

RL ARCH. VIROL. 140:1473-1482(1995).
DR EMBL; D38543; D1008127; -.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00767; Poly_coat; 1.
DR PFAM; PF00863; Peptidase_C4; 1.
KW POLYPROTEIN; COAT PROTEIN.
FT NON_TER 1
FT CHAIN 1 16 CYTOPLASMIC INCLUSION PROTEIN.
FT CHAIN 17 69 6K2 PROTEIN.
FT CHAIN 70 261 VPG PROTEIN.
FT CHAIN 262 504 NIA PROTEASE.
FT CHAIN 505 1025 NUCLEAR INCLUSION PROTEIN B.
FT CHAIN 1026 1340 COAT PROTEIN.
SQ SEQUENCE 1340 AA; 151873 MW; 12C01867 CRC32;

Query Match 59.3%; Score 51; DB 14; Length 1340;
Best Local Similarity 50.0%; Pred. No. 5.58e+00;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 146 DPLTGAVIDDSPYT 159
   |||||:|
QY 66 DPSTGALVDSKSYA 79

RESULT 9
ID O39734 PRELIMINARY; PRT; 3493 AA.
AC O39734;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POLYPROTEIN.
OS SWEET POTATO FEATHERY MOTTLE VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC POTYVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S STRAIN;
RA NISHIGUCHI M.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S STRAIN;
RX MEDLINE; 98336489.
RA SAKAI J., MORI M., MORISHITA T., TANAKA M., HANADA K., USUGI T.,
RA NISHIGUCHI M.;
RT "Complete nucleotide sequence and genome organization of sweet potato
RT feathery mottle virus (S strain) genomic RNA: the large coding region
RT of the P1 gene.";
RL ARCH. VIROL. 142:1553-1562(1997).
DR EMBL; D86371; D1023570; -.
DR PFAM; PF00271; helicase_C; 1.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00767; Poly_coat; 1.
DR PFAM; PF00851; Peptidase_C6; 1.
DR PFAM; PF00863; Peptidase_C4; 1.
KW POLYPROTEIN.
FT CHAIN 1 664 POTENTIAL.
FT CHAIN 665 1122 POTENTIAL.
FT CHAIN 1123 1474 POTENTIAL.
FT CHAIN 1475 1526 POTENTIAL.
FT CHAIN 1527 2169 POTENTIAL.
FT CHAIN 2170 2222 POTENTIAL.
FT CHAIN 2223 2414 GENOME-LINKED VIRAL PROTEIN.
FT CHAIN 2415 2657 PROTEASE.
FT CHAIN 2658 3178 REPLICASE.
FT CHAIN 3179 3493 COAT PROTEIN.
SQ SEQUENCE 3493 AA; 393818 MW; A02EFA05 CRC32;

Query Match 59.3%; Score 51; DB 14; Length 3493;
Best Local Similarity 50.0%; Pred. No. 5.58e+00;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 2299 DPLTGAVIDDSPYT 2312

```



```

ID Q46094 PRELIMINARY; PRT; 368 AA.
AC Q46094;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HEAT SHOCK PROTEIN/SERINE PROTEASE (FRAGMENT).
GN HTRA.
OS CAMPYLOBACTER JEJUNI.
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; CAMPYLOBACTER GROUP;
OC CAMPYLOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UA580.
RX MEDLINE; 90384493.
RA TAYLOR D.E., HIRATSUKA K.;
RT "Use of non-radioactive DNA probes for detection of Campylobacter
RT jejuni and Campylobacter coli in stool specimens.";
RL MOL. CELL. PROBES 4:261-271(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-UA580.
RA HIRATSUKA K.;
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U27271; G881375; -.
DR PFAM; PF00089; trypsin; 1.
DR PFAM; PF00595; PDZ; 2.
KW HEAT SHOCK; PROTEASE.
FT NON_TER 1
SQ SEQUENCE 368 AA; 39491 MW; A5062589 CRC32;

Query Match 65.1%; Score 56; DB 2; Length 368;
Best Local Similarity 72.7%; Pred. No. 4.07e-01;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 121 SGGALVDSRGY 131
QY 68 STGALVDSKSY 78

RESULT 3
ID Q46120 PRELIMINARY; PRT; 472 AA.
AC Q46120;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SERINE PROTEASE.
GN HTRA.
OS CAMPYLOBACTER JEJUNI.
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; CAMPYLOBACTER GROUP;
OC CAMPYLOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-81116;
RA HENDERSON J., WOOD A.C., EMERY M.J., WREN B.W., KETLEY J.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X82628; E315368; -.
DR PFAM; PF00089; trypsin; 1.
DR PFAM; PF00595; PDZ; 2.
KW PROTEASE.
SQ SEQUENCE 472 AA; 50940 MW; 6CBCA101 CRC32;

Query Match 65.1%; Score 56; DB 2; Length 472;
Best Local Similarity 72.7%; Pred. No. 4.07e-01;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 225 SGGALVDSRGY 235
QY 68 STGALVDSKSY 78

RESULT 4
ID O60320 PRELIMINARY; PRT; 405 AA.
AC O60320;

```

```

DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE KIAA0574 PROTEIN (FRAGMENT).
GN KIAA0574.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN.
RX MEDLINE; 98290545.
RA NAGASE T., ISHIKAWA K., MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N.,
RA OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT code for large proteins in vitro.";
RL DNA RES. 5:31-39(1998).
DR EMBL; AB011146; D1026430; -.
FT NON_TER 1
SQ SEQUENCE 405 AA; 42409 MW; 09BA9932 CRC32;

Query Match 60.5%; Score 52; DB 4; Length 405;
Best Local Similarity 77.8%; Pred. No. 3.35e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 345 SLVDSKAYA 353
QY 71 ALVDSKSYA 79

RESULT 5
ID O46178 PRELIMINARY; PRT; 552 AA.
AC O46178;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE RADIAL SPOKEHEAD.
OS STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; EUECHINOIDEA;
OC ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE; STRONGYLOCENTROTUS.
RN [1]
RP SEQUENCE FROM N.A.
RX GINGRAS D., GAGNON C.;
RT "Molecular cloning and characterization of a radial spoke head
RT protein of sea urchin sperm axonemes: involvement of the protein in
RT the regulation of sperm motility.";
RL MOL. BIOL. CELL 9:513-522(1998).
DR EMBL; U73123; G2905895; -.
SQ SEQUENCE 552 AA; 62723 MW; 898CFCCC CRC32;

Query Match 60.5%; Score 52; DB 5; Length 552;
Best Local Similarity 50.0%; Pred. No. 3.35e+00;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 39 PTEALVNAKAY 50
QY 67 PSTGALVDSKSY 78

RESULT 6
ID O60309 PRELIMINARY; PRT; 870 AA.
AC O60309;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE KIAA0563 PROTEIN.
GN KIAA0563.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]

```

\*\*\*\*\*  
W A I T E S R A H  
\*\*\*\*\*  
(TW)

Release 3.1A John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_Pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:29:24 1999; MasPar time 5.26 Seconds  
Tabular output not generated. 145.242 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (66-79) from PCTUS9913024.pep (11 of 12)  
Perfect Score: 86  
Sequence: 1 DPSTGALVDSKSYA 14

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sprembl9  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 24.488; Variance 26.991; scale 0.907

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	59	68.6	1577	2 Q54178	GLUCOSYLTRANSFERASE	7.85e-02
2	56	65.1	368	2 Q46094	HEAT SHOCK PROTEIN/SER	4.07e-01
3	56	65.1	472	2 Q46120	SERINE PROTEASE	4.07e-01
4	52	60.5	405	4 O60320	KIAA0574 PROTEIN (FRAG	3.35e-00
5	52	60.5	552	5 O46178	RADIAL SPOKEHEAD	3.35e-00
6	52	60.5	870	4 O60309	KIAA0563 PROTEIN	3.35e-00
7	51	59.3	540	5 O24281	PHOTOLYASE	5.58e-00
8	51	59.3	1340	14 Q82874	POLYPROTEIN (FRAGMENT)	5.58e-00
9	51	59.3	3493	14 Q39734	POLYPROTEIN	5.58e-00
10	50	58.1	142	14 Q87600	ENVELOPE GLYCOPROTEIN	9.22e+00
11	50	58.1	156	11 O35308	MONOCARBOXYLATE TRANSP	9.22e+00
12	50	58.1	262	2 Q44188	W-AMINO-TRANSFERASE-LI	9.22e+00
13	50	58.1	371	2 P72097	CMP-N-ACETYLNEURAMINAT	9.22e+00
14	50	58.1	492	11 O70461	MONOCARBOXYLATE TRANSP	9.22e+00
15	50	58.1	542	13 O3151	RETINAL EPITHELIAL TRANSP	9.22e+00
16	50	58.1	542	13 Q30632	MONOCARBOXYLATE TRANSP	9.22e+00
17	50	58.1	1133	10 O04626	SIMILARITY TO MYOSIN H	9.22e+00
18	50	58.1	1378	11 Q61138	PATERNALLY EXPRESSED P	9.22e+00
19	49	57.0	319	2 O34966	YCDH	1.51e+01
20	49	57.0	327	14 Q92324	HEMAGGLUTININ (FRAGMENT)	1.51e+01

21	49	57.0	329	14	O40846	HEMAGGLUTININ (FRAGMENT)	1.51e+01
22	49	57.0	329	14	O40849	HEMAGGLUTININ (FRAGMENT)	1.51e+01
23	49	57.0	427	5	Q25991	PARASITOPHOROUS VACUOL	1.51e+01
24	49	57.0	734	11	O88970	INSULIN RECEPTOR SUBST	1.51e+01
25	49	57.0	966	1	O26770	HYPOTHETICAL 104.9 KD	1.51e+01
26	49	57.0	984	5	Q26016	SERINE RICH PROTEIN (S	1.51e+01
27	49	57.0	1139	2	Q54073	ANCHOR PROTEIN, LCM	1.51e+01
28	48	55.8	195	2	O06554	HYPOTHETICAL 20.8 KD P	2.46e+01
29	48	55.8	328	14	O82581	HAEMAGGLUTININ (HA1 DO	2.46e+01
30	48	55.8	329	14	O40888	HEMAGGLUTININ (FRAGMENT)	2.46e+01
31	48	55.8	329	14	O40887	HEMAGGLUTININ (FRAGMENT)	2.46e+01
32	48	55.8	329	14	O70688	HEMAGGLUTININ (FRAGMENT)	2.46e+01
33	48	55.8	329	14	O40901	HEMAGGLUTININ (FRAGMENT)	2.46e+01
34	48	55.8	329	14	O70682	HEMAGGLUTININ (FRAGMENT)	2.46e+01
35	48	55.8	329	14	O70684	HEMAGGLUTININ (FRAGMENT)	2.46e+01
36	48	55.8	427	10	Q03349	S-LOCUS-SPECIFIC GLYCO	2.46e+01
37	48	55.8	477	11	O70142	SCK, PARTIAL CDS (FRAG	2.46e+01
38	48	55.8	480	2	O88047	PUTATIVE TRANSCRIPTION	2.46e+01
39	48	55.8	540	4	O60230	SCK, PARTIAL CDS (FRAG	2.46e+01
40	48	55.8	550	14	O82753	HAEMAGGLUTININ PRECURS	2.46e+01
41	48	55.8	580	3	O13902	DIHYDROXYACETONE KINAS	2.46e+01
42	48	55.8	712	2	O46520	OMPID PRECURSOR	2.46e+01
43	48	55.8	1571	11	O54978	ZINC FINGER PROTEIN	2.46e+01
44	47	54.7	329	14	O40654	HEMAGGLUTININ (FRAGMENT)	3.97e+01
45	47	54.7	329	14	O67138	HEMAGGLUTININ (FRAGMENT)	3.97e+01

ALIGNMENTS

RESULT 1  
ID Q54178 PRELIMINARY; PRT: 1577 AA.  
AC Q54178; Q54247;  
DT 01-NOV-1996 (TREMBLER. 01, CREATED)  
DT 01-NOV-1996 (TREMBLER. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLER. 08, LAST ANNOTATION UPDATE)  
DE GLUCOSYLTRANSFERASE.  
GN GTFG.  
OS STREPTOCOCCUS GORDONII.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
OC STREPTOCOCCUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CHALLIS;  
RX MEDLINE; 96157084.  
RA VICKERMAN M.M., SULAVIK M.C., CLEWELL D.B.;  
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase  
phase variants";  
RL DEV. BIOL. STAND. 85:309-314(1995).  
RN [2]  
RP SEQUENCE OF 1-96 FROM N.A.  
RC STRAIN-CHALLIS;  
RX MEDLINE; 92276337.  
RA SULAVIK M.C., TARDIF G., CLEWELL D.B.;  
RT "Identification of a gene, rgg, which regulates expression of  
glucosyltransferase and influences the spp phenotype of Streptococcus  
gordonii Challis";  
RL J. BACTERIOL. 174:3577-3586(1992).  
DR EMBL; U12643; G1054877; -;  
DR EMBL; M89776; G153795; -;  
DR PFAM; PF00128; alpha-amylase; 1.  
KW TRANSFERASE.  
SQ SEQUENCE 1577 AA; 177805 MW; 7D58FEC2 CRC32;  
  
Query Match 68.6%; Score 59; DB 2; Length 1577;  
Best Local Similarity 69.2%; Pred. No. 7.85e-02;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
Db 260 DAETGALVDSNEY 272  
|: |||||:  
QY 66 DPSTGALVDSKSY 78

RESULT 2

[illegible]

RP	SEQUENCE FROM N.A.					
RA	HARRIS B.;					
RL	SUBMITTED (DEC-1996)	TO EMBL/GENBANK/DDBJ DATA BANKS.				
RN	[2]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE; 94150718.					
RA	WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,					
RA	BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,					
RA	CRAFTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,					
RA	GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,					
RA	JONES M., KERSHAW J., KRISTEN J., LAISTER N., LATREILLE P.,					
RA	LIGHTNING J., LLOYD C., MCURRAY A., MORTIMORE B., O'CALLAGHAN M.,					
RA	PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOMKEEN R.,					
RA	SHALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,					
RA	THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,					
RA	WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;					
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.					
RT	elegans";					
RL	NATURE 368:32-38(1994).					
DR	EMBL; 283220; E1344511; -					
SQ	SEQUENCE 921 AA; 106698 MW; 21DE4C35 CRC32;					
Query Match 15.2%; Score 84; DB 5; Length 921;						
Best Local Similarity 24.7%; Pred. No.1.21e+00;						
Matches 19; Conservative 22; Mismatches 32; Indels 4; Gaps						
Db	212 ERNFVEIVHGVRQEYFIPLIG-RISLTIGRRSTKYACTRFLKRGNPTGVNAVYVETE 270					
QY	::   :     :: : : : : : : : : : : : : : : : : : : :					
Db	271 QI-VMDMASSGNVDGR 286					
QY	: :   :   :   :   :   :   :   :   :   :   :   :   :					
QY	61 HLRVFD-PSTGALVDSK 76					
RESULT 12						
ID	O49984	PRELIMINARY;	PRT;	180 AA.		
AC	O49984;					
DT	01-JUN-1998 (TREMBREL. 06, CREATED)					
DT	01-JUN-1998 (TREMBREL. 06, LAST SEQUENCE UPDATE)					
DT	01-JUN-1998 (TREMBREL. 06, LAST ANNOTATION UPDATE)					
DE	PUTATIVE ETHYLENE RECEPTOR (FRAGMENT).					
OS	BOERS.					
GN	BRASSICA OLERACEA (CAULIFLOWER).					
OC	EUKARYOTA; VIRIDIPANTAE; STRIPTOPHYTA; EMERYOPHYTA; TRACHEOPHYTA;					
OC	EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;					
OC	CAPPALES; BRASSICACEAE; BRASSICA.					
[1]						
RP	SEQUENCE FROM N.A.					
RA	CHARNG Y.Y., SUN C.W., YAN S.L., CHOU S.J., CHEN Y.R., YANG S.F.;					
RL	SUBMITTED (JAN-1997)	TO EMBL/GENBANK/DDBJ DATA BANKS.				
DR	EMBL; U87239; G2738025; -					
FT	NON_TER 1					
FT	NON_TER 180 180					
SQ	SEQUENCE 180 AA; 19859 MW; D9E072C1 CRC32;					
Query Match 15.1%; Score 83; DB 10; Length 180;						
Best Local Similarity 36.0%; Pred. No.1.68e+00;						
Matches 18; Conservative 9; Mismatches 20; Indels 3; Gaps						
Db	48 IPTCTCLAKPGVPVKAFPEYVSVRVL-LHLSNFGQSDMSDL-SGKGY 95					
QY	: :      :    :   : : : : : : : : : : : : : : : : : :					
QY	30 IAGSCPLSTAGPSYVKFQDNPVGSQTFSAGLHLRVDFDST-CALVDSKSY 78					
RESULT 13						
ID	Q38022	PRELIMINARY;	PRT;	281 AA.		
AC	Q38022;					
DT	01-NOV-1996 (TREMBREL. 01, CREATED)					
DT	01-NOV-1996 (TREMBREL. 01, LAST SEQUENCE UPDATE)					
DT	01-NOV-1996 (TREMBREL. 08, LAST ANNOTATION UPDATE)					
DE	ORF 1.					
OS	BACTERIOPHAGE PHI-C31.					

```
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE HEPATITIS E VIRUS COMPLETE GENOME.
OS HEPATITIS E VIRUS (HEV).
OC VIRUSES: SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;
OC CALICIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92024067.
RA TAM A.W., SMITH M.M., GUERRA M.E., HUANG C.C., BRADLEY D.W., FRY K.E.,
RA REYES G.R.;
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT full-length viral genome.";
RL VIROLOGY 185:120-131(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92261377.
RA UCHIDA T., SUZUKI K., HAYASHI N., IIDA F., HARA T., OO S.S.,
RA WANG C.K., SHIKATA T., ICHIKAWA M., RIKIHISA T., MIZUNO K., WIN K.M.;
RT "Hepatitis E virus: cDNA cloning and expression.";
RL MICROBIOL. IMMUNOL. 36:67-79(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92335008.
RA AYE T.T., UCHIDA T., MA X.Z., IIDA F., SHIKATA T., ZHUANG H.,
RA WIN K.M.;
RT "Complete nucleotide sequence of a hepatitis E virus isolated from
RT the Xinjiang epidemic (1986-1988) of China.";
RL NUCLEIC ACIDS RES. 20:3512-3512(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92115700.
RA TSAREV S.A., EMERSON S.U., REYES G.R., TSAREVA T.S., LECTERS L.J.,
RA MALIK I.A., IQBAL M., PURCELL R.H.;
RT "Characterization of a prototype strain of hepatitis E virus.";
RL PROC. NATL. ACAD. SCI. U.S.A. 89:559-563(1992).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92271462.
RA FRY K.E., TAM A.W., SMITH M.M., KIM J.P., LUK K.C., YOUNG L.M.,
RA PIATAK M., FELDMAN R.A., YUN K.Y., PURDY M.A., ET AL.;
RT "Hepatitis E virus (HEV): strain variation in the nonstructural gene
RT region encoding consensus motifs for an RNA-dependent RNA polymerase
RT and an ATP/GTP binding site.";
RL VIRUS GENES 6:173-185(1992).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93348763.
RA BI S.L., PURDY M.A., MCCAUSTLAND K.A., MARGOLIS H.S., BRADLEY D.W.;
RT "The sequence of hepatitis E virus isolated directly from a single
RT source during an outbreak in China.";
RL VIRUS RES. 28:233-247(1993).
DR EMBL: L08816; G33009; -.
SQ SEQUENCE 1693 AA; 185052 MW; 30A917E1 CRC32;

Query Match 15.6%; Score 86; DB 14; Length 1693;
Best Local Similarity 25.4%; Pred. No. 6.27e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 FRTSFVDGAVLENGPERHNLSPDASQSTMAAGPFSLTYYAASAGLEVRYVAAGLDHRAV 631
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 8 FGQGYVOTPFLESNSVRYKISIAAGSCPLSTAGPSYKVFQDNPVGSQT-F-SAGLHLR-V 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 632 FAPGVSP 638
| : : :
QY 65 FDPSTGA 71

RESULT 7
ID Q81344 PRELIMINARY; PRT; 1693 AA.
AC Q81344;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

Db 572 FRTSFVDGAVLENGPERHNLSPDASQSTMAAGPFSLTYYAASAGLEVRYVAAGLDHRAV 631
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 8 FGQGYVOTPFLESNSVRYKISIAAGSCPLSTAGPSYKVFQDNPVGSQT-F-SAGLHLR-V 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 632 FAPGVSP 638
| : : :
QY 65 FDPSTGA 71

RESULT 8
ID Q89444 PRELIMINARY; PRT; 1693 AA.
AC Q89444;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE UNNAMED PROTEIN PRODUCT.
OS HEPATITIS E VIRUS (HEV).
OC VIRUSES: SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;
OC CALICIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RA YIN S.R., PURCELL R.H., EMERSON S.U.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K52-87;
RX MEDLINE: 95176571.
RA YIN S., PURCELL R.H., EMERSON S.U.;
RT "A new Chinese isolate of hepatitis E virus: comparison with strains
RT recovered from different geographical regions.";
RL VIRUS GENES 9:23-32(1994).
DR EMBL: L25547; G1209366; -.
DR EMBL: L25595; G784878; -.
SQ SEQUENCE 1693 AA; 185122 MW; 0E56663A CRC32;

Query Match 15.6%; Score 86; DB 14; Length 1693;
Best Local Similarity 25.4%; Pred. No. 6.27e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 FRTSFVDGAVLENGPERHNLSPDASQSTMAAGPFSLTYYAASAGLEVRYVAAGLDHRAV 631
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 8 FGQGYVOTPFLESNSVRYKISIAAGSCPLSTAGPSYKVFQDNPVGSQT-F-SAGLHLR-V 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 632 FAPGVSP 638
| : : :
QY 65 FDPSTGA 71

RESULT 9
ID Q69410 PRELIMINARY; PRT; 1693 AA.
AC Q69410;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
```

```
RA LI Y., LU Z., SUN L., ROPP S., KUTISH G.F., ROCK D.L.,
RA VAN ETEN J.L.;
RT "Analysis of 74 kb of DNA located at the right end of the 330-kb
RT chloroella virus PCV-1 genome.";
RL VIROLOGY 237:360-377(1997).
DR EMBL: U42580; G2447164;
SQ SEQUENCE 114 AA; 13257 MW; 1B75BE43 CRC32;

Query Match 17.4%; Score 96; DB 14; Length 114;
Best Local Similarity 27.4%; Pred. No. 2,00e-02;
Matches 17; Conservative 18; Mismatches 25; Indels 2; Gaps 1;

Db 5 F5MSNISRVLRTASCAI--LNPSYGRKSRNNSLFTMSLSLWVFVSIKPSMSK 62
QY | | : : : : : : : : : : : : : : : : : : : : : : : : : :
| | : : : : : : : : : : : : : : : : : : : : : : : : : :
17 FLSESNVRYKISAGSCPLSTAGPSYKVFQDNVPVGSQTF-SAGLHLRVPDPTGALVDSK 76

Db 63 KY 64
QY 77 SY 78

RESULT 3
ID Q65574 PRELIMINARY; PRT; 299 AA.
AC Q65574;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE HYPOTHETICAL 32.4 KD PROTEIN.
GN UL7.
OS BOVINE HERPESVIRUS TYPE 1.
OC VIRUSES: DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAPERESVIRINAE; VARICELLOVIRUS.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-SCHOENBOEKEN;
RX MEDLINE: 9513343.
RA VLCEK C., BENES V., LU Z., KUTISH G.F., PACES V., ROCK D.,
RA LETCHWORTH G.J., SCHWYZER M.;
RT "Nucleotide sequence analysis of a 30-kb region of the bovine
RT herpesvirus 1 genome which exhibits a collinear gene arrangement with
RT the UL21 to UL4 genes of herpes simplex virus.";
RL VIROLOGY 210:100-108(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-JURA;
RA SCHWYZER M.;
RA LOWERY D.E., SIMARD C., BELLO L.J., THIRY E., VLCEK C.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-JURA;
RA SCHWYZER M.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBBJ DATA BANKS.
DR EMBL: 248053; G971327;
DR EMBL: A7004801; E1187337;
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 299 AA; 32379 MW; A3897D4A CRC32;

Query Match 16.7%; Score 92; DB 14; Length 299;
Best Local Similarity 47.2%; Pred. No. 8.15e-02;
Matches 17; Conservative 9; Mismatches 6; Indels 4; Gaps 4;

Db 45 PRFVCEVREIPAGPPTFTSSSITHLRV-EPSTGALL 79
QY | | : : : : : : : : : : : : : : : : : : : : : : : : : :
| | : : : : : : : : : : : : : : : : : : : : : : : : : :
41 PSYV-KFQDNVPVGSQTF-SAGL-HLRVPDPTGALV 73

RESULT 4
ID Q65581 PRELIMINARY; PRT; 300 AA.
AC Q65581;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
```

```
DE UL7 POLYPEPTIDE.
GN UL7.
OS BOVINE HERPESVIRUS TYPE 1.
OC VIRUSES: DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAPERESVIRINAE; VARICELLOVIRUS.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-SCHOENBOEKEN;
RX MEDLINE: 96135223.
RA SCHWITT J., KEIL G.M.;
RT "Identification and characterization of the bovine herpesvirus 1 UL7
RT gene and gene product which are not essential for virus replication
RT in cell culture.";
RL J. VIROL. 70:1091-1099(1996).
DR EMBL: X91751; G1006630;
SQ SEQUENCE 300 AA; 32450 MW; 86377347 CRC32;

Query Match 16.7%; Score 92; DB 14; Length 300;
Best Local Similarity 47.2%; Pred. No. 8.15e-02;
Matches 17; Conservative 9; Mismatches 6; Indels 4; Gaps 4;

Db 46 PRFVCEVREIPAGPPTFTSSSITHLRV-EPSTGALL 80
QY | | : : : : : : : : : : : : : : : : : : : : : : : : : :
| | : : : : : : : : : : : : : : : : : : : : : : : : : :
41 PSYV-KFQDNVPVGSQTF-SAGL-HLRVPDPTGALV 73

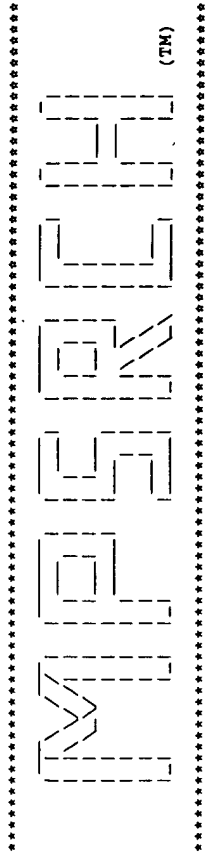
RESULT 5
ID O84354 PRELIMINARY; PRT; 566 AA.
AC O84354;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 63.5 KD PROTEIN.
GN CT350.
OS CHLAMYDIA TRACHOMATIS.
OC BACTERIA; CHLAMYDIALES; CHLAMYDIACEAE; CHLAMYDIA.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-D/UW-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATISOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT Chlamydia trachomatis.";
RL SCIENCE 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-D/UW-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATISOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBBJ DATA BANKS.
DR EMBL: AE001308; G3328771;
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 566 AA; 63508 MW; C1A1C491 CRC32;

Query Match 16.2%; Score 89; DB 2; Length 566;
Best Local Similarity 26.2%; Pred. No. 2.28e-01;
Matches 17; Conservative 25; Mismatches 18; Indels 5; Gaps 5;

Db 63 TAESYLOQSFISEDTYIR-KSAIIGA-GLSGSSEA-LELLSEAIETDLYEQL-L-ILNA 117
QY : : : : : : : : : : : : : : : : : : : : : : : : : :
| | : : : : : : : : : : : : : : : : : : : : : : : : : :
8 FGQGYVQTFPLESNVRYKISAGSCPLSTAGPSYKVFQDNVPVGSQTF-SAGLHLRVPDP 67

Db 118 ATSQL 122
QY : : : : : : : : : : : : : : : : : : : : : : : : : :
| | : : : : : : : : : : : : : : : : : : : : : : : : : :
68 STGAL 72

RESULT 6
ID O81862 PRELIMINARY; PRT; 1693 AA.
AC O81862;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
```



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Sep 1 16:13:05 1999; MasPar time 9.45 Seconds  
Tabular output not generated. 450.572 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (1-78) from PCTUS9913024.pep (5 of 12)  
Sequence: 1 MEKFAEFGQGVQTFPLSE.....GLHLRVFDPSTGALVDSKY 78

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl9  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 37.051; Variance 67.149; scale 0.552

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	100	18.1	1693	14	COMPLETE GENOME SEQUEN	4.78e-03
2	96	17.4	114	14	A606L PROTEIN.	2.00e-02
3	92	16.7	299	14	HYPOTHETICAL 32.4 KD P	8.15e-02
4	92	16.7	300	14	UL7 POLYPEPTIDE.	8.15e-02
5	89	16.2	566	2	HYPOTHETICAL 63.5 KD P	2.28e-01
6	86	15.6	1693	14	ORF 1.	6.27e-01
7	86	15.6	1693	14	HEPATITIS E VIRUS COM	6.27e-01
8	86	15.6	1693	14	UNNAMED PROTEIN PRODUC	6.27e-01
9	86	15.6	1693	14	METHYL TRANSFERASE.	6.27e-01
10	84	15.2	622	2	CRYIIC DELTA-ENDOTOXIN	1.21e+00
11	84	15.2	921	5	C34B7.2 PROTEIN.	1.21e+00
12	83	15.1	180	10	PUTATIVE ETHYLENE RECE	1.68e+00
13	83	15.1	281	9	ORF 1.	1.68e+00
14	83	15.1	1577	2	GLUCOSYLTRANSFERASE.	1.68e+00
15	83	15.1	1693	14	NONSTRUCTURAL POLYPROT	1.68e+00
16	81	14.7	196	14	PUTATIVE COAT PROTEIN.	3.21e+00
17	81	14.7	216	2	FLAGELLAR L-RING PROTE	3.21e+00
18	81	14.7	633	2	INSECTICIDAL CRYSTAL P	3.21e+00
19	81	14.7	642	14	PUTATIVE READTHROUGH P	3.21e+00
20	81	14.7	3011	14	GENOME POLYPROTEIN.	3.21e+00

21	80	14.5	275	14	Q68467	POLYPROTEIN (FRAGMENT)	4.41e+00
22	80	14.5	275	14	Q68468	POLYPROTEIN (FRAGMENT)	4.41e+00
23	80	14.5	275	14	Q68469	POLYPROTEIN (FRAGMENT)	4.41e+00
24	80	14.5	357	5	Q01993	VC27A7L.1 PROTEIN	4.41e+00
25	80	14.5	802	3	Q07034	RNA BINDING PROTEIN.	4.41e+00
26	80	14.5	2436	14	Q81756	POLYPROTEIN (FRAGMENT)	4.41e+00
27	79	14.3	370	3	Q04083	D9461.24P.	6.06e+00
28	79	14.3	542	5	Q17456	SIMILAR TO GLUTAMATE D	6.06e+00
29	79	14.3	860	1	Q59003	860AA LONG HYPOTHETICA	6.06e+00
30	78	14.2	70	2	P75688	FROM BASES 311709 TO 3	8.28e+00
31	78	14.2	613	10	Q38846	ETHYLENE RESPONSE SENS	8.28e+00
32	78	14.2	3011	14	Q36609	POLYPROTEIN.	8.28e+00
33	78	14.2	3011	14	Q36608	POLYPROTEIN.	8.28e+00
34	78	14.2	3011	14	Q36610	POLYPROTEIN.	8.28e+00
35	78	14.2	3011	14	Q36579	POLYPROTEIN.	8.28e+00
36	77	14.0	418	2	Q56631	LECITHINASE.	1.13e+01
37	77	14.0	470	2	O87325	LECITHINASE.	1.13e+01
38	77	14.0	510	10	O65322	PUTATIVE MONOSACCHARID	1.13e+01
39	77	14.0	967	2	Q54123	PUBP.	1.13e+01
40	77	14.0	1361	2	O86617	HYPOTHETICAL 145.0 KD	1.13e+01
41	77	14.0	1464	11	O08948	N-METHYL-D-ASPARTATE R	1.13e+01
42	77	14.0	1464	11	O63728	N-METHYL-D-ASPARTATE R	1.13e+01
43	77	14.0	1464	4	Q12879	N-METHYL-D-ASPARTATE R	1.13e+01
44	77	14.0	2219	5	Q23388	ZK1067.2 PROTEIN.	1.13e+01
45	76	13.8	821	2	O51735	OUTER MEMBRANE PROTEIN	1.54e+01

ALIGNMENTS

RESULT 1  
ID Q81876 PRELIMINARY; PRT: 1693 AA.  
AC Q81876;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)  
DE COMPLETE GENOME SEQUENCE.  
OS HEPATITIS E VIRUS (HEV).  
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALCIVIRIDAE;  
RN CALICIVIRUS.  
RC SEQUENCE FROM N.A.  
RA UCHIDA T.;  
RL SUBMITTED (MAY-1992) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: D11093; D1002342;  
SQ SEQUENCE 1693 AA; 184840 MW; 8952DD38 CRC32;

Query Match 18.1%; Score 100; DB 14; Length 1693;  
Best Local Similarity 26.9%; Pred. No. 4.78e-03;  
Matches 18; Conservative 22; Mismatches 24; Indels 3; Gaps 3;

Db	572	FRTSFVDCAVLEANGPERYNLSFDASQSTMAAGPFLSYAASAGLEVRYVAAGLDHRAV	631
Qy	8	FGQGVQTFPLSESNVRYKISAGSCPLSTAGPSYVFKQDNPVGSQT-F-SAGLHLR-V	64
Db	632	FAPGVSP	638
Qy	65	FDPSTGA	71

RESULT 2  
ID Q41088 PRELIMINARY; PRT: 114 AA.  
AC Q41088;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE A606L PROTEIN.  
OS PARAMECIDIUM BURSARIA CHLORELLA VIRUS 1 (PBCV-1).  
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PHYCODNAVIRIDAE; PHYCODNAVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98022962.

```
CC  -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC -----
CC  EMBL: Z36098; G536526; -
CC  PIR: S46105; S46105.
CC  PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
CC  PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; FALSE_NEG.
CC  PFAM: PF01055; Glycosyl_hydrl5; 1.
CC  KW  HYPOTHETICAL PROTEIN: HYDROLASE; GLYCOSIDASE.
CC  FT  ACT_SITE 537 BY SIMILARITY.
CC  SQ  SEQUENCE 954 AA; 110265 MW; 9710BA62 CRC32;
CC -----
CC  Query Match 14.5%; Score 80; DB 1; Length 954;
CC  Best Local Similarity 21.5%; Pred. No. 1.59e+00;
CC  Matches 14; Conservative 20; Mismatches 29; Indels 2; Gaps 2;
CC -----
Db 136 KFOEENRNTSIPQHFLLKQQTVNSFWSKISSFLSLNSTADTFHLRNGDYSVEIFAEPF 195
QY 3 KFAAEFGQGVV-QTPFLSESNVR-YKISAGSCPLSTAGPSYVKFQDNPVGSQTSAGL 60
Db 196 QLKVV 200
QY 61 HLRVF 65
Search completed: Wed Sep 1 16:12:47 1999
Job time : 15 secs.
```





```
FT CONFLICT 45 45 H -> R (IN REF. 2).
SQ SEQUENCE 527 AA; 61239 MW; 8DF019E3 CRC32;

Query Match
  14.9% Score 82; DB 1; Length 527;
Best Local Similarity 44.4%; Pred. No. 7.96e-01;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 242 LRVPGMGSCPLPATGFSYKWEIDP 268
   ||::|||::|||::|||::|||::|||
QY 24 VRYKISAGSCPLSTAGPSYKFDNP 50

RESULT 11
ID TAX_MOUSE STANDARD; PRT; 527 AA.
AC P42682;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE TXK (EC 2.7.1.112) (PTK-RL-18) (RESTING
DN LYMPHOCYTE KINASE).
GN TXK OR RLK.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=THYMUS;
RX MEDLINE; 96059536.
RA HAIRE R.N., LITMAN G.W.;
RT "The murine form of TXK, a novel TEC kinase expressed in thymus maps
RT to chromosome 5.",
RL MAMM. GENOME 6:476-480(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=THYMUS;
RA SOMMERS C.L., HUANG K., GRINBERG A., CHARLICK D.A., KOZAK C.A.,
RA LOVE P.E.;
RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=LIVER;
RA HIGASHITSUJI H., NONOGUCHI K., ARII S., FURUTANI M., KANEKO Y.,
RA NAKAYAMA H., FUJITA J.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=THYMUS;
RX MEDLINE; 95130578.
RA HU O., DAVIDSON D., SCHWARTZBERG P.L., MACCHIARINI F.,
RA LENARDO M.J., BLUESTONE J.A., MATIS L.A.;
RT "Identification of Rlk, a novel protein tyrosine kinase with
RT predominant expression in the T cell lineage.",
RL J. BIOL. CHEM. 270:1928-1934(1995).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN EARLY THYMOCYTES, T CELLS AND
CC MAST CELLS.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE BTK SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U16145; G562125; -;
DR EMBL; U19607; G643065; -;
```

```
DR EMBL; D43963; G604884; -;
DR EMBL; L35268; G623443; -;
DR MGD; MGI:102960; TXK.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00018; SH2; 1.
DR PFAM; PF00017; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSP; Q06187; IAWW.
KW TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;
KW SH3 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 14 20 POLY-CYS.
FT DOMAIN 82 142 SH3.
FT DOMAIN 150 246 SH2.
FT DOMAIN 271 527 PROTEIN KINASE.
FT NP_BIND 277 285 ATP (BY SIMILARITY).
FT BINDING 299 299 ATP (BY SIMILARITY).
FT ACT_SITE 390 390 BY SIMILARITY.
FT MOD_RES 420 420 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 3 4 LS -> SF (IN REF. 3).
FT CONFLICT 6 6 Y -> D (IN REF. 3).
FT CONFLICT 272 272 A -> T (IN REF. 3).
FT CONFLICT 497 497 R -> S (IN REF. 3).
SQ SEQUENCE 527 AA; 61108 MW; 5B39DA78 CRC32;

Query Match
  14.7% Score 81; DB 1; Length 527;
Best Local Similarity 44.4%; Pred. No. 1.13e+00;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 242 LRVPIGLGSCPLPATGFSYKWEIDP 268
   ||::|||::|||::|||::|||::|||
QY 24 VRYKISAGSCPLSTAGPSYKFDNP 50

RESULT 12
ID YCB2_PSEDE STANDARD; PRT; 171 AA.
AC P29944;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 19.0 KD PROTEIN IN COBS 5'REGION (ORF2).
OS PSEUDOMONAS DENITRIFICANS.
OC BACTERIA; PROTEOBACTERIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92011364.
RA CAMERON B., GUILHOT C., BLANCHE F., CAUCHOIS L., ROUYEZ M.-C.,
RA RIGAUD S., LEVY-SCHIL S., CROUZET J.;
RT "Genetic and sequence analyses of a Pseudomonas denitrificans DNA
RT fragment containing two cob genes.",
RL J. BACTERIOL. 173:6058-6065(1991).
CC -!- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M62869; G151167; -;
DR PIR; B38162; B38162.
DR PROSITE; PS00636; DNAJ_1; FALSE_NEG.
DR PROSITE; PS50076; DNAJ_2; 1.
DR PFAM; PF00226; DnaJ; 1.
DR HSP; P08622; IXL.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 171 AA; 18973 MW; 5DDDD21D5 CRC32;
```



```

CC EMBL; M80581; G329998;
KW POLYPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; HELICASE;
FT NP-BINDING. 975 982 ATP (POTENTIAL).
SQ SEQUENCE 1693 AA; 185149 MW; FBCA2483 CRC32;

Query Match 15.6%; Score 86; DB 1; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.92e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 FRTSVDGAVLETNGPERHNLSFASQSTMAAGPFSLTYSAAAGLEVVRYVAAGLDHRAV 631
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 8 FGQYVQTPFLSESNVRYKISAGSPLSTAGPSYVKFQDNPVGSQT-F-SAGLHLR-V 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 632 FAPGVSP 638
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 FDPSTGA 71

RESULT 6
ID Y4II_RHISN STANDARD; PRT; 703 AA.
AC P55492;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 76.2 KD PROTEIN Y4II.
GN Y4II.
OS RHIZOBIUM SP. (STRAIN NGR234).
OG PLASMID SYN PNGR234A.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; RHIZOBIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97305956.
RA FREIBERG C.A., FELLAY R., BAIROCH A., BROUGHTON W.J., ROSENTHAL A.,
RA PERRET X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL NATURE 387:394-401(1997).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: NONE OBVIOUS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000078; G2182446;
KW HYPOTHETICAL PROTEIN; PLASMID; TRANSMEMBRANE.
FT TRANSMEM 23 43 POTENTIAL.
FT TRANSMEM 69 89 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 250 270 POTENTIAL.
FT TRANSMEM 357 377 POTENTIAL.
FT TRANSMEM 432 452 POTENTIAL.
FT TRANSMEM 644 664 POTENTIAL.
SQ SEQUENCE 703 AA; 76183 MW; A2BA53CE CRC32;

Query Match 15.4%; Score 85; DB 1; Length 703;
Best Local Similarity 27.0%; Pred. No. 2.75e-01;
Matches 17; Conservative 17; Mismatches 28; Indels 1; Gaps 1;

Db 224 SPADTDSGTSIATSLFGAGASTRPEWLAFAANALAGAPPIGLSPRVATATSOYA 283
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 15 TPLF-SESNVRYKISAGSPLSTAGPSYVKFQDNPVGSQTFSGAGLHLRVPDPTGALV 73
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 KAK 286
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 74 DSK 76
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 7
ID CPAC_ECOLI STANDARD; PRT; 869 AA.
AC P25733;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE CFA/I FIMBRIAL SUBUNIT C PRECURSOR (COLONISATION FACTOR ANTIGEN I
DE SUBUNIT C).
GN CPAC.
OS ESCHERICHIA COLI.
OG PLASMID NTP513.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ENTEROTOXIGENIC;
RX MEDLINE; 89330163.
RA HAMERS A.M., PEL H.J., WILLSHAW G.A., KUSTERS J.G.,
RA VAN DER ZEIJST B.A.M., GAASTRA W.;
RT "The nucleotide sequence of the first two genes of the CFA/I fimbrial
RT operon of human enterotoxigenic Escherichia coli.";
RL MICROB. PATHOG. 6:297-309(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92329981.
RA JORDI B.J.A.M., WILLSHAW G.A., VAN DER ZEIJST B.A.M., GAASTRA W.;
RT "The complete nucleotide sequence of region 1 of the CFA/I fimbrial
RT operon of human enterotoxigenic Escherichia coli.";
RL DNA SEQ. 2:257-263(1992).
CC -1- FUNCTION: MAY SERVE AS ANCHOR FOR THE FIMBRIAE IN THE OUTER
CC MEMBRANE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M55661; G145510;
KW ANTIGEN; SIGNAL; FIMBRIA; OUTER MEMBRANE; PLASMID.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 869 CFA/I FIMBRIAL SUBUNIT C.
SQ SEQUENCE 869 AA; 97830 MW; 7AF76347 CRC32;

Query Match 15.4%; Score 85; DB 1; Length 869;
Best Local Similarity 32.4%; Pred. No. 2.75e-01;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;

Db 158 AFTQSQTINLSDCKYKRLSISGNSALGITDTSY 191
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 11 GYVQTPFLSESNVRYK-ISIAGSPLSTAGPSY 43
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
ID YX45_MYCTU STANDARD; PRT; 611 AA.
AC Q10830;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 67.2 KD PROTEIN CY274.45C.
GN MYCY274.45C OR MYCY338.01C.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FRMICTUTES; ACTINOBACTERIA; ACTINOBACTERIACEAE; MYCOBACTERIUM.
OC ACTINOMYCETALES; CORYNEBACTERIACEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RA CONNOR R., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: BELONGS TO THE N-ACYL-D-AMINO-ACID DEACYLASE FAMILY.

```

```
CC -----
CC TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10330; G221705; -
CC POLYPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; HELICASE;
CC ATP-BINDING.
CC NP_BIND 975 982 ATP (POTENTIAL).
CC SEQUENCE 1693 AA; 185215 MW; FFCB786D CRC32;
CC -----
Query Match 18.1%; Score 100; DB 1; Length 1693;
Best Local Similarity 26.9%; Pred. No. 9.43e-04;
Matches 18; Conservative 22; Mismatches 24; Indels 3; Gaps 3;
CC -----
Db 572 FRTSFVDGAVLEANGPERYNLSFDASQSTMAAGPFSLTAAAGLEVRVYVAAAGLDHRAV 631
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 8 FGQGYVQTPFLSESNVRYKISAGSCPLSTAGSPYKVDNPNVGSQT-F-SAGHLR-V 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 632 FAGVSP 638
| : : :
QY 65 FDPSTGA 71
| : : :
CC -----
RESULT 3
ID EMB8_PICGL STANDARD; PRT; 457 AA.
AC Q40863;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE LATE EMBRYOGENESIS ABUNDANT PROTEIN EMB8.
GN EMB8.
OS PICEA GLAUCA (WHITE SPRUCE).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; CONIFEROPSIDA; CONIFERALES; PINACEAE;
OC PICEA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L47118; G1350545; -
CC PROSITE; PS01193; UPF0017; 1.
CC SEQUENCE 457 AA; 51019 MW; AE7CB4CD CRC32;
CC -----
Query Match 15.6%; Score 86; DB 1; Length 457;
Best Local Similarity 34.0%; Pred. No. 1.92e-01;
Matches 17; Conservative 11; Mismatches 20; Indels 2; Gaps 2;
CC -----
Db 232 LGANILRYLGEVAGNCPLSGAVSLNPF-NLVIADDFHKGGLGFNNVYD 280
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 18 LSESNVRYKISAGSCPLSTAGSPYKVDNPNVGSQTFSSAGHLR-VFD 66
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
CC -----
RESULT 4
ID POLN_HEVBU STANDARD; PRT; 1693 AA.
AC P29324;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
```

```
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE
DE (EC 2.7.7.48); HELICASE].
OS HEPATITIS E VIRUS (STRAIN BURMA) (HEV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;
OC CALICIVIRUS.
CC -----
CC SEQUENCE FROM N.A.
CC MEDLINE; 92024067.
CC TAM A.W., SMITH M.M., GUERRA M.E., HUANG C.-C., BRADLEY D.W.,
CC FRY K.E., REYES G.R.;
CC "Hepatitis E virus (HEV): molecular cloning and sequencing of the
CC full-length viral genome.";
CC VIROLOGY 185:120-131(1991).
CC -I- HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF ENTERICALLY
CC TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M73218; G330024; -
CC PIR; A40778; MNWHE.
CC POLYPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; HELICASE;
CC ATP-BINDING.
CC NP_BIND 975 982 ATP (POTENTIAL).
CC SEQUENCE 1693 AA; 185191 MW; C560BE14 CRC32;
CC -----
Query Match 15.6%; Score 86; DB 1; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.92e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;
CC -----
Db 572 FRTSFVDGAVLENGPERHNLSDASQSTMAAGPFSLTAAAGLEVRVYVAAAGLDHRAV 631
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 8 FGQGYVQTPFLSESNVRYKISAGSCPLSTAGSPYKVDNPNVGSQT-F-SAGHLR-V 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 632 FAGVSP 638
| : : :
QY 65 FDPSTGA 71
| : : :
CC -----
RESULT 5
ID POLN_HEVPA STANDARD; PRT; 1693 AA.
AC P33424;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE
DE (EC 2.7.7.48); HELICASE].
OS HEPATITIS E VIRUS (STRAIN PAKISTAN) (HEV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;
OC CALICIVIRUS.
CC -----
CC SEQUENCE FROM N.A.
CC MEDLINE; 92115700.
CC TSAREV S.A., EMERSON S.U., REYES G.R., TSAREVA T.S., LECTERS L.J.,
CC MALIK I.A., IQBAL M., PURCELL R.H.;
CC "Characterization of a prototype strain of hepatitis E virus.";
CC PROC. NATL. ACAD. SCI. U.S.A. 89:559-563(1992).
CC -I- HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF ENTERICALLY
CC TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
```



```

Best Local Similarity 44.4%; Pred.No. 2.96e+00;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 242 LRYPIGLLGSCLPATSGFSYKWEIDP 268
   || ||: ||| :|| || | :|
Qy 24 VRYKISAGSCPLSTAGPSYVKQDNP 50

RESULT 14
ENTRY Txk - mouse
TITLE #type complete
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
12-Feb-1999

ACCESSIONS I49133
REFERENCE I49133
#authors Haire, R.N.; Litman, G.W.
#journal Mann. Genome (1995) 6:476-480
#title The murine form of TXK, a novel TEC kinase expressed in
#cross-references MUID:96059536
#accession I49133
#status preliminary; translated from GB/EMBL/DDBJ
#molecule_type mRNA
#residues 1-527 #label RES
#cross-references EMBL:U016145; NID:g562124; PID:g562125
CLASSIFICATION #superfamily protein-tyrosine kinase tec; pleckstrin repeat
homology; protein kinase homology; SH2 homology; SH3
homology
KEYWORDS ATP
FEATURE
89-137 #domain SH3 homology #label SH3\
150-246 #domain SH2 homology #label SH2\
269-527 #domain protein kinase homology #label KIN\
277-285 #region protein kinase ATP-binding motif
SUMMARY #length 527 #molecular-weight 61108 #checksum 785

Query Match 14.7%; Score 81; DB 2; Length 527;
Best Local Similarity 44.4%; Pred.No. 2.96e+00;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 242 LRYPIGLLGSCLPATSGFSYKWEIDP 268
   || ||: ||| :|| || | :|
Qy 24 VRYKISAGSCPLSTAGPSYVKQDNP 50

RESULT 15
ENTRY S40770 #type complete
TITLE polyprotein precursor - hepatitis C virus
ORGANISM #formal_name hepatitis C virus
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
18-Sep-1998

ACCESSIONS S40770; PC1285
REFERENCE S40770
#authors Okamoto, H.
#submission submitted to the EMBL Data Library, March 1992
#accession S40770
#status preliminary
#molecule_type genomic RNA
#residues 1-3011 #label OKA
#cross-references EMBL:D10749; NID:g221586; PID:d1002057; PID:g221587
PC1284
#authors Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka,
T.; Yoshizawa, H.; Tsuda, F.; Miyakawa, Y.; Mayumi, M.
#journal J. Exp. Med. (1990) 60:167-177
#title The 5'-terminal sequence of the hepatitis C virus genome.
#accession PC1285
#molecule_type genomic RNA
#residues 1-513 #label OK2
#experimental_source isolate HC-J1
CLASSIFICATION #superfamily hepatitis C virus genome polyprotein; DEAD/H box
helicase homology
polyprotein
KEYWORDS

```





```
ALTERNATE_NAMES      plasmid
delta-endotoxin
#formal_name Bacillus thuringiensis
#sequence_revision 10-Nov-1995 #text_change
DATE 22-Nov-1993
ACCESSIONS           S17400
REFERENCE            S17402
#authors             Wu, D.; Cao, X.L.; Bai, Y.Y.; Aronson, A.I.
#journal             FEMS Microbiol. Lett. (1991) 81:31-36
#title               Sequence of an operon containing a novel delta-endotoxin gene
                    from Bacillus thuringiensis.
#accession           S17402
##status             preliminary
##molecule_type     DNA
##residues           1-622 #label WUD
##cross-references   EMBL:X57252; NID:g40283; PID:g40286
GENETICS
#genome              plasmid
#keywords             delta-endotoxin
SUMMARY
#length 622 #molecular-weight 69729 #checksum 8205
Query Match          15.2%; Score 84; DB 2; Length 622;
Best Local Similarity 18.2%; Pred. No. 1.13e+00;
Matches 12; Conservative 26; Mismatches 26; Indels 2; Gaps 2;
Db 501 FISEKYNQDLSRFELSNPTARYTLRGNNSYNLRLVSSIGSSIRVTINGRVYTNV 560
QY 12 YVOTPFLESNSVRYKISAGS-CPLSTAGPSYVKF-QDNPVGSQTFSGAGLHLRVDPST 69
Db 561 NTTTNN 566
QY 70 GALVDS 75

RESULT 7
ENTRY S38913 #type complete
TITLE hypothetical protein 1 - phage phi-C31
ORGANISM #formal_name phage phi-C31
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
ACCESSIONS S38913
REFERENCE S38912
#authors Hartley, N.M.; Murphy, G.O.; Bruton, C.J.; Chater, K.F.
#submission submitted to the EMBL Data Library, November 1993
#accession S38913
##status preliminary
##molecule_type     DNA
##residues           1-281 #label HAR
##cross-references   EMBL:X76288; NID:g432610; PID:g579071
GENETICS
#start_codon        GTG
SUMMARY
#length 281 #molecular-weight 31680 #checksum 2790
Query Match          15.1%; Score 83; DB 2; Length 281;
Best Local Similarity 24.1%; Pred. No. 1.56e+00;
Matches 14; Conservative 17; Mismatches 24; Indels 3; Gaps 3;
Db 194 AYMNADFIIDPGNRPMPFEFGAAVLHYVTETWA-FKPVTGPDVFAQFLHLRQTFD 250
QY 11 GYVQTPFLESNSVRYKIS-IAGSCPLSTAGPSYVKFQDNPVGSQTFSGAGLHLR-VFD 66

RESULT 8
ENTRY D70928 #type complete
TITLE hypothetical protein Rv2913c - Mycobacterium tuberculosis
                    (strain H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
ACCESSIONS D70928
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                    C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry
```

```
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
                    the complete genome sequence.
#cross-references MUID:98295987
#accession D70928
##status preliminary; nucleic acid sequence not shown;
                    translation not shown
##molecule_type     DNA
##residues           1-611 #label COL
##cross-references   GB:274024; GB:AL123456; NID:g3250700; PID:el301028;
                    PID:g3250707
##experimental_source strain H37Rv
GENETICS
#gene Rv2913c
#summary #length 611 #molecular-weight 67204 #checksum 9971
Query Match          15.1%; Score 83; DB 2; Length 611;
Best Local Similarity 43.5%; Pred. No. 1.56e+00;
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
Db 319 VRFOHLPVPFELYSGDIDLVPVE 341
QY 44 VKFQDNPVGSQTFSGAGLHLRVFD 66

RESULT 9
ENTRY I38375 #type fragment
TITLE tyrosine kinase - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 29-May-1998 #sequence_revision 29-May-1998 #text_change
ACCESSIONS I38375
REFERENCE I38372
#authors Haire, R.N.; Ohta, Y.; Lewis, J.E.; Fu, S.M.; Kroisel, P.;
                    Litman, G.W.
#journal Hum. Mol. Genet. (1994) 3:897-901
#title TXK, a novel human tyrosine kinase expressed in T cells
                    shares sequence identity with Tec family kinases and maps
                    to 4p12.
#cross-references MUID:95038742
#accession I38375
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type     DNA
##residues           1-81 #label RES
##cross-references   EMBL:U07794; NID:g508219; PID:g508224
GENETICS
#gene GDB:TXK
#cross-references   GDB:377329; OMIM:600058
#map_position 4p12-4p12
#introns 25/1
CLASSIFICATION
#superfamily protein-tyrosine kinase tec; pleckstrin repeat
                    homology; protein kinase homology; SH2 homology; SH3
                    homology
FEATURE
32-81 #domain protein kinase homology (fragment) #label KIN
SUMMARY
#length 81 #checksum 7751
Query Match          14.9%; Score 82; DB 2; Length 81;
Best Local Similarity 44.4%; Pred. No. 2.15e+00;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Db 5 LRPVGLMGSLPAPAGSYKWEIDP 31
QY 24 VRYKISAGSCPLSTAGPSYVKFQDNP 50
```

```
#cross-references MUID:99000809
#accession G71525
##status preliminary
##molecule_type DNA
##residues 1-566 #label ARN
##cross-references GB:AE001308; GB:AE001273; NID:g3328766; PID:g3328771
##experimental_source serotype D, strain UW-3/Cx
GENETICS
#gene CT350
#length 566 #molecular-weight 63507 #checksum 4960
SUMMARY
Query Match 16.2%; Score 89; DB 2; Length 566;
Best Local Similarity 26.2%; Pred. No. 2.14e-01;
Matches 17; Conservative 25; Mismatches 18; Indels 5; Gaps 5;

Db 63 IAEYLSQSFSEDTYIR-KSAIGA-GLSGSSEA-LELLSEATQDLYEQL-L-ILNA 117
QY 8 FGQGYVQTPFLSESNSVRYKISIAAGSCLPLTAGPSYKVFQDNPVGSQTFSAGLHLRVDP 67
Db 118 ATSQL 122
QY 68 STGAL 72

RESULT 3
ENTRY MNWHE #type complete
TITLE genome polyprotein - hepatitis E virus (strain Burma)
CONTAINS RNA-directed RNA polymerase (EC 2.7.7.48)
ORGANISM #formal_name hepatitis E virus
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
ACCESSIONS A40778; A48547
REFERENCE A40778
#authors Tam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, G.R.
#journal Virology (1991) 185:120-131
#title Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome.
#cross-references MUID:92024067
#accession A40778
##residues 1-1693 #label TAM
##molecule_type genomic RNA
##cross-references GB:M73218; NID:g330023; PID:g330024
REFERENCE A48547
#authors Fry, K.E.; Tam, A.W.; Smith, M.M.; Kim, J.P.; Luk, K.C.; Young, L.M.; Piatlak, M.; Feldman, R.A.; Yun, K.Y.; Purdy, M.A.; McCaustland, K.A.; Bradley, D.W.; Reyes, G.R.
#journal Virus Genes (1992) 6:173-185
#title Hepatitis E virus (HEV): strain variation in the nonstructural gene region encoding consensus motifs for an RNA-dependent RNA polymerase and an ATP/GTP binding site.
#cross-references MUID:92271462
#accession A48547
##molecule_type genomic RNA
##residues 967-1693 #label FRY
##cross-references GB:M32400; NID:g330021; PID:g330022
##note sequence extracted from NCBI backbone (NCBIN:104572, NCBI:P104573)
CLASSIFICATION #superfamily hepatitis E virus nonstructural protein
KEYWORDS ATP; nonstructural protein; nucleotidyltransferase
SUMMARY #length 1693 #molecular-weight 185191 #checksum 6520

Query Match 15.6%; Score 86; DB 1; Length 1693;
Best Local Similarity 25.4%; Pred. No. 5.84e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 FRTSFVDGAVLENGPERNLSDASQSTWAGPFFSLTYAASAGLEVRYVAGLDRHV 631
QY 8 FGQGYVQTPFLSESNSVRYKISIAAGSCLPLTAGPSYKVFQDNPVGSQT-F-SAGLHLR-V 64
Db 632 FAPGVSP 638
QY 65 FDPSTGA 71
```

```
RESULT 4
ENTRY A61046 #type complete
TITLE ecdysonine-induced membrane protein IMP-E3 - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Feb-1997
ACCESSIONS A61046
REFERENCE A61046
#authors Moore, J.T.; Fristrom, D.; Hammonds, A.S.; Fristrom, J.W.
#journal Dev. Genet. (1990) 11:299-309
#title Characterization of IMP-E3, a gene active during imaginal disc morphogenesis in Drosophila melanogaster.
#accession A61046
##status preliminary
##molecule_type mRNA
##residues 1-331 #label MOO
GENETICS
#gene FlyBase:ImpE3
##cross-references FlyBase:FBgn0001255
KEYWORDS membrane protein
SUMMARY #length 331 #molecular-weight 36583 #checksum 8221

Query Match 15.4%; Score 85; DB 2; Length 331;
Best Local Similarity 31.9%; Pred. No. 8.12e-01;
Matches 15; Conservative 12; Mismatches 18; Indels 2; Gaps 2;

Db 190 LDNFLRLYDNTGRAAFSGESAMDRWSTASIAGKKRPPTKP-YVDF 235
QY 1 MEKFMAEFGQGVQVQTPFLSESNSVRYKI-SIAGSCLPLTAGPSYKVF 46

RESULT 5
ENTRY C56617 #type complete
TITLE cfac protein precursor - Escherichia coli plasmid NTP113
ORGANISM #formal_name Escherichia coli
DATE 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 20-Mar-1998
ACCESSIONS C56617
REFERENCE A56617
#authors Jorði, B.J.; Willshaw, G.A.; van der Zeijst, B.A.; Gaastra, W.
#journal DNA Seq. (1992) 2:257-263
#title The complete nucleotide sequence of region 1 of the CFA/I fimbrial operon of human enterotoxigenic Escherichia coli.
#cross-references MUID:92329981
#accession C56617
##status preliminary
##molecule_type DNA
##residues 1-869 #label JOR
##cross-references GB:M55661; NID:g145507; PID:g145510
##experimental_source enterotoxigenic strain, CFA/I-St plasmid NTP113
##note sequence extracted from NCBI backbone (NCBIN:108960, NCBI:P108971)
GENETICS
#gene cfac
#genome plasmid
SUMMARY #length 869 #molecular-weight 97830 #checksum 9755

Query Match 15.4%; Score 85; DB 2; Length 869;
Best Local Similarity 32.4%; Pred. No. 8.12e-01;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;

Db 158 AFIQSQTINLSDSGYKRLISGNSALGITDTSY 191
QY 11 GYVQTPFLSESNSVRYK-ISIAGSCLPLTAGPSY 43

RESULT 6
ENTRY S17402 #type complete
TITLE parasporal crystal protein cryIIc - Bacillus thuringiensis
```

Result No.	Score	Query		DB	ID	Description	Pred. No.
		Match	Length				
1	92	16.7	299	2	G61248	hypothetical protein	7.69e-02
2	89	16.2	566	2	S71525	probable OMP [leader	2.14e-01
3	86	15.6	1093	1	MNWHHE	genome polyprotein -	5.84e-01
4	85	15.4	331	2	A61046	ecdysone-induced memb	8.12e-01
5	85	15.4	869	2	C56617	cfcAC protein precursor	8.12e-01
6	84	15.2	622	2	S17402	parasporal crystal pr	1.13e+00
7	83	15.1	281	2	S38913	hypothetical protein	1.56e+00
8	83	15.1	611	2	D70928	hypothetical protein	1.56e+00
9	82	14.9	81	2	I38375	tyrosine kinase - hum	2.15e+00
10	82	14.9	449	2	S5092	hypothetical protein	2.15e+00
11	82	14.9	527	2	I84483	tyrosine kinase - hum	2.15e+00
12	81	14.7	216	2	G70447	flagellar L-ring prot	2.96e+00
13	81	14.7	527	2	A56331	protein-tyrosine kina	2.96e+00
14	81	14.7	527	2	I49133	Ttk - mouse	2.96e+00
15	81	14.7	3011	2	S40770	polyprotein precursor	2.96e+00
16	80	14.5	171	2	B38162	hypothetical protein	4.05e+00
17	80	14.5	572	2	S14200	GRESAC protein - Tryp	4.05e+00
18	80	14.5	802	2	S48529	NAB3 protein - yeast	4.05e+00
19	80	14.5	954	2	S46105	glucan 1,4-alpha-gluc	4.05e+00
20	80	14.5	1659	2	JC4956	vitellogenin precursor	4.05e+00
21	80	14.5	3011	1	GNWVC3	genome polyprotein -	4.05e+00
22	79	14.3	370	2	S69718	hypothetical protein	5.55e+00
23	79	14.3	472	1	A35322	vitamin D-binding pro	5.55e+00

Search completed: Wed Sep 1 16:15:06 1999  
Job time : 37 secs.

HCV genomic amino acid sequence isolated from infected human LG.  
Hepatitis C Virus; Non-A, non-B hepatitis virus; HCV; NANBHV;  
human growth hormone; HGH; secretion signal; fusion protein;

Hepatitis C Virus.

M09315193-A.

PD 05-AUG-1993.

PJ 29-JAN-1993; UO0907.

PR 31-JAN-1992; US-830024..

PA (ABBO ) ABBOTT LAB.

PI Bode SL, Casey JM., Desai SM, Devare SG, Fraile DE;

PI Yamaguchi J, Zeck BJ;

DR WPI : 93-258673/32.

PT New plasmid pHCv-162 is a mammalian expression systems for HCVD  
proteins - useful for diagnosing HCV infection and as vaccines  
for preventing HCV infection

PS Example 1 ; Page 39-49; 100pp; English.

CC RNA was isolated from the plasma of a HCV seropositive human  
(designated "LG") and cDNA was prepared from it. The cDNA was  
PCR amplified using specific primers with sequences based  
on the prototype HCV-1 cDNA sequence (GENBANK M62321). Further  
amplification using nested primers resulted in 7 adjacent HCV DNA  
fragments which could be assembled into a full-length sequence. The  
cDNA sequence was determined and translated into the genomic amino  
acid sequence. Comparison of the LG genomic amino acid sequence  
with that from HCV-1 showed 134 amino acid differences.

SQ Sequence 3011 AA;

Query Match 14.7% Score 81; DB 8; Length 3011;  
Best Local Similarity 40.0%; Pred. No. 2.72e+01;  
Matches 18; Conservative 10; Mismatches 13; Indels 4; Gaps 4;

Dbb 2116 vpspeffeldgvrhlr-rfappckpllrddevsfvgldhydvgsq 2159  
| |::|| ::|| ::|| ::||  
Y q 13 VQTPEFUSENSRYRIKSIAGSCPLSTAGPSPVKFDONPVGSG 54

```

RESULT 14
ID P92033 standard; protein; 170 AA
AC P92033
DT 02-MAR-1990 (first entry)
DE Sequence encoded in the Hepatitis C virus (HCV) CDNA insert in clone 14C
KW Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH).
OS Hepatitis C virus
FH Key_ Location/Qualifiers
FT region 1..71
   /*tag= a
EP-318216-A.
PN 31-MAY-1989.
PD
PF 18-NOV-1988; 310922.
PR 14-NOV-1988; US-271450; US-122714.
PA (CHIR) Chiron Corp.
PI Houghton M, Choo Q-L, Kuo G;
DR WPI; 89-159274/22.
DR N-PSDB; N92089.
PT Purified hepatitis C virus
   and associated nucleic acids and polypeptide(s)
PT Claim 13; Figure 18; 139pp; English.
FS CC It is the sequence encoded in the Hepatitis C virus (HCV) CDNA insert in
CC clone 14C. Tag a = the region of overlap with the HCV antigen encoded in
CC clone 25C. It is an epitope which could be used as immunoassay reagents
CC and vaccines and to generate antibodies useful in diagnosis and passive
CC immunotherapy for HCV infection/non-A, non-B hepatitis.
CC Sequence 170 AA;

Query Match 14.5%; Score 80; DB 1; Length 170;
Best Local Similarity 37.88; Pred. No. 3.26e+01;
Matches 17; Conservative 11; Mismatches 13; Indels 4; Gaps 4;

Db 63 vspseffldgvrll-rfappckpllreevsfvrlghyvpvgq 106
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 13 VQRP-FLSESNVSRYKISTAGSC-FLSTAGPSY-VKFODNPVGSO 54

```

```
DR WPI: 98-506734/43.  
DR N-PSDB; V52611.  
PT New insecticidal Bacillus thuringiensis toxins - useful for  
PT controlling lepidopteran pests, especially Ostrinia nubilalis,  
PT Heliothis virescens and Helicoverpa zea  
PS Claim 14; Pages 32-34; 50pp; English.  
CC This is the amino acid sequence of a novel Bacillus thuringiensis toxin  
CC used in the method of the invention, to control lepidopteran pests.  
CC The new toxins are useful as pesticides, especially for the control of  
CC Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The  
CC polynucleotide coding sequences are useful for recombinant expression  
CC of the toxins and the primers, together with probes derived from the  
CC new sequences, are useful for the identification and characterisation  
CC of novel genes that encode pesticidal toxins.  
SQ Sequence 633 AA;  
  
Query Match          15.4%; Score 85; DB 35; Length 633;  
Best Local Similarity 23.7%; Pred.No.1.31e+01;  
Matches 14; Conservative 20; Mismatches 23; Indels 2; Gaps 2  
  
Db      51l fisekfngdsirfegsnrtarytlrgrnsgnylnrvssignstvirtingryvtas 569  
QY    : : |::|::|: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:  
12 YVQTFFLSESNVRYKIS-TAGSCPLSTAGPSYVKF-QDNPVGSOTFSAGHLRLVFDPDS 68  
  
RESULT 10  
ID RI4618 standard; Protein; 1693 AA.  
AC RI4618;  
DT 16-JAN-1992 (first entry)  
DE Protein encoded by ORF 1 of Burmese ET-NANB viral strain.  
KW enterically transmitted non-A, non-B hepatitis virus; Hepatitis C;  
KC HCV; E.coli strain B84; ATCC deposit number 67717; Burma.  
OS Enterically transmitted non-A, non-B hepatitis virus.  
PN W09115603-A.  
PD 17-OCT-1991.  
PF 05-APR-1991; U02358.  
PR 03-APR-1990; US-505880.  
PA (GENE-) GENELABS INC.  
PP ((US)) ) US DEPT HEALTH & HUMAN.  
PI Reyes GR, Yarbough PO, Bradley DW, Krawczynski KZ, Tam A;  
PI FTY KE;  
DR WFI: 91-325242/44.  
DR N-PSDB; Q14412.  
PT New viral proteins from non A-non-B hepatitis agent - used to  
PT treat and prevent enterically-transmitted non-A non-B hepatitis  
PT virus  
PS Disclosure; Page 15; 117pp; English.  
CC A positive clone EtI.1 was identified in a library prepared from  
CC bile of cynomolgus monkeys infected with the Burma strain of ET-NANB.  
CC Both strands of EtI.1 were sequenced. Identity of the sequence with  
CC sequences in etiologic agents has been confirmed by locating a  
CC similar sequence in a viral strain isolated in Burma. This protein  
CC is encoded by the longest ORF (ORF 1) of the Burma strain.  
CC (See Q14410 for EtI.1).  
SQ Sequence 1693 AA;  
  
Query Match          15.2%; Score 84; DB 3; Length 1693;  
Best Local Similarity 26.6%; Pred.No.1.57e+01;  
Matches 17; Conservative 20; Mismatches 24; Indels 3; Gaps 3  
  
Db      572 frtsfdvgvletngperhnlsifdasgntnaagfsityaasaagletryvaagldhrav 631  
QY    | ::|: | ::|: | ::|: | ::|: | ::|: | ::|: | ::|: | ::|: |  
8 FGQGYVTPTFLSESNSVRYKIISTAGCPLSTAGPSYVKFDONPVGSOI-F-SAGHLRL-V 64  
  
Db      632 fapq 635  
QY    | ::|: | ::|: | ::|: | ::|: | ::|: | ::|: | ::|: | ::|: |  
        65 FDPDS 68  
  
RESULT 11  
ID W75775 standard; Protein; 633 AA.  
AC W75775;  
DT 02-DEC-1998 (first entry)
```



AC W76368; 1998 (first entry)  
 DT 03-DEC-1998 (first entry)  
 DE Hepatitis E virus hollow particle protein #1.  
 KW Hepatitis E virus hollow particle protein; virus; antibody; detection; immunoassay;  
 OS Infection.  
 OS Hepatitis virus.  
 FH Key  
 FT Location/Qualifiers  
 FT 1..1693  
 FT /note= "Partial sequence"  
 PN J10234383-A.  
 PN 08-SEP-1998.  
 PD 28-FEB-1997; 062445.  
 PR 28-FEB-1997; JP-062445.  
 PA (DENK-) DENKA SEIKEN KK.  
 PA (KOKU-) KOKURITSU YOBO EISEI KENKYUSHO.  
 DR WPI: 98-535037/46.  
 DR N-PSDB: V61687.  
 PT Hepatitis E virus hollow particle poly:peptide(s) and nucleic acids  
 PT encoding it - useful for more accurate detection of HEV in samples,  
 PT using immuno-assays and nucleic acid hybridisation  
 PS Claim 10; Page 17-24; 29pp; Japanese.  
 CC This sequence represents a Hepatitis E viral hollow particle protein.  
 CC This polypeptides can be used to raise antibodies to detect HEV  
 CC infection in samples, e.g. by immuno-assay based techniques, and the  
 CC nucleic acid can be used for the same in nucleic acid hybridisation  
 CC assays. The polypeptides and nucleic acids allow more accurate  
 CC detection of HEV than previously possible.  
 SQ Sequence 1693 AA;  
 Query Match 17.4%; Score 96; DB 35; Length 1693;  
 Best Local Similarity 26.9%; Pred. No. 1.66e+00;  
 Matches 18; Conservative 21; Mismatches 25; Indels 3; Gaps 3;  
 Db 572 frtsfdvgavleangperynlfsdasgstmaagpfsptyaasaglevryvaagldhrav 631  
 QY 8 FGQGYVQTPLSESNVRYKISAGSCPLSTAGPSYVKFQDNPVGSOT-F-SAGLHLR-V 64  
 Db 632 fapgvsp 638  
 QY 65 FDPSTGA 71  
 RESULT 3  
 ID W75773 standard; Protein; 623 AA.  
 AC W75773;  
 DT 02-DEC-1998 (first entry)  
 DE Amino acid sequence of lepidoteran-active HD573 toxin.  
 KW HD573 toxin; PCR; primer; amplification; Bacillus thuringiensis; probe;  
 KW lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;  
 KW Helicoverpa zea; hybridisation.  
 OS Bacillus thuringiensis.  
 PN WO9840490-A1.  
 PD 17-SEP-1998.  
 PF 13-MAR-1998; U05081.  
 PR 13-MAR-1997; US-040512.  
 PA (MYCO ) MYCOGEN CORP.  
 PI Muller-Cohn J, Narva KE, Schnepf HE;  
 DR WPI: 98-506734/43.  
 DR N-PSDB: V52610.  
 PT New insecticidal Bacillus thuringiensis toxins - useful for  
 PT controlling lepidopteran pests, especially Ostrinia nubilalis,  
 PT Heliothis virescens and Helicoverpa zea  
 PS Claim 14; Pages 28-30; 50pp; English.  
 CC This is the amino acid sequence of a novel Bacillus thuringiensis toxin  
 CC used in the method of the invention, to control lepidopteran pests.  
 CC The new toxins are useful as pesticides, especially for the control of  
 CC Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The  
 CC polynucleotide coding sequences are useful for recombinant expression  
 CC of the toxins and the primers, together with probes derived from the  
 CC new sequences, are useful for the identification and characterisation  
 CC of novel genes that encode pesticidal toxins.  
 SQ Sequence 623 AA;

Query Match 15.8%; Score 87; DB 35; Length 623;  
 Best Local Similarity 18.2%; Pred. No. 9.05e+00;  
 Matches 12; Conservative 26; Mismatches 26; Indels 2; Gaps 2;  
 Db 502 fisekyngqdsirfelsesntartylrgnsgnsnylrvsigsstirvtingrvytanv 561  
 QY 12 YVOTPFLESNSVRYKIS-IGASCPSTAGPSYVKF-QDNPVGSQTF-SAGLHLR-VFDPST 69  
 Db 562 ntntnn 567  
 QY 70 GALVDS 75  
 RESULT 4  
 ID R51264 standard; Protein; 1693 AA.  
 AC R51264;  
 DT 21-OCT-1994 (first entry)  
 DE HEV strain protein encoded by ORF-1.  
 KW Hepatitis E virus; HEV; strain SAR-55; open reading frame; ORF;  
 KW antibody; detection; diagnosis; primates; stool suspension.  
 OS Hepatitis E virus strain SAR-55.  
 PN WO9406913-A.  
 PD 31-MAR-1994.  
 PF 17-SEP-1993; U08849.  
 PR 18-SEP-1992; US-947263.  
 PA (USSH ) US SEC DEPT HEALTH.  
 PI Emerson SU, Purcell RH, Tsarev SA;  
 DR WPI: 94-118462/14.  
 DR N-PSDB: Q45197.  
 PT Purified hepatitis E strain SAR-55 virus - used to develop prods.  
 PT for use in detection, diagnosis, vaccines and therapy of  
 PT hepatitis E virus infection  
 PS Disclosure: Page 57-62; 114pp; English.  
 CC The sequences given in R51264-66 are encoded by the hepatitis E virus  
 CC (HEV) strain SAR-55. The cDNA sequence contains three open reading  
 CC frames (ORFs). These proteins can be used to stimulate the production  
 CC of protective antibodies upon injection into a mammal that would serve  
 CC to protect the mammal upon challenge with wild type HEV. The proteins  
 CC can be used for detection and diagnosis of HEV infection. The HEV  
 CC SAR-55 cDNA was isolated from primates inoculated with stool  
 CC suspensions obtained from hepatitis E patients.  
 SQ Sequence 1693 AA;  
 Query Match 15.6%; Score 86; DB 10; Length 1693;  
 Best Local Similarity 25.4%; Pred. No. 1.09e+01;  
 Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;  
 Db 572 frtsfdvgavleangperhnlfsdasgstmaagpfsptyaasaglevryvaagldhrav 631  
 QY 8 FGQGYVQTPLSESNVRYKISAGSCPLSTAGPSYVKFQDNPVGSOT-F-SAGLHLR-V 64  
 Db 632 fapgvsp 638  
 QY 65 FDPSTGA 71  
 RESULT 5  
 ID W80196 standard; Protein; 1693 AA.  
 AC W80196;  
 DT 23-DEC-1998 (first entry)  
 DE Protein encoded by ORF1 of ET-NANB (HEV) Burma strain DNA sequence.  
 KW Enterically transmitted non A non B hepatitis virus; ET-NANB;  
 KW Hepatitis E virus; HEV; Burma HEV isolate; vaccine;  
 KW diagnostic probe  
 OS Non A non B Hepatitis virus.  
 PN US5824649-A.  
 PD 20-OCT-1998.  
 PF 07-JUN-1995; 475807.  
 PR 25-JUL-1994; US-279823.  
 PR 17-JUN-1988; US-208997.  
 PR 11-APR-1989; US-336672.  
 PR 16-JUN-1989; US-367486.  
 PR 13-OCT-1989; US-420921.





Search completed: Wed Sep 1 16:10:34 1999  
Job time : 24 secs.



```

[2]
RN EMBL: D37935; D1007731; -
RP RC STRAIN-BRISTOL N2;
RA DU Z., GAVTUNG S.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U50312; G1226312; -
DR PFAM: PF00282; Pyridoxal_dec; 1.
SQ SEQUENCE 542 AA; 61133 MW; A7C148AB CRC32;

Query Match 19.7%; Score 79; DB 5; Length 542;
Best Local Similarity 30.3%; Pred. No. 2.10e+00;
Matches 10; Conservative 11; Mismatches 10; Indels 2; Gaps 2;

Db 309 GFLP-PFLEE-DEIRYDFRVPGVSSISADSHKY 339
      ||| | : || : || : || : || : ||
QY 11 GYVQTPFLSESNSVRYKISAGSCPLSTAGPSY 43

RESULT 9
ID O59003 PRELIMINARY; PRT; 860 AA.
AC O59003;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE 860AA LONG HYPOTHETICAL PROTEIN.
GN PH1302.
OS PYROCOCUS HORIKOSHII.
OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCCLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OT3;
RX MEDLINE; 98344137.
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOIYAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT *Complete Sequence and Gene Organization of the Genome of a
RT Hyper-thermophilic Archaeobacterium, Pyrococcus horikoshii OT3.*;
RL DNA RES. 5:55-76(1998)
DR EMBL; AF000006; D1031349; -
SQ SEQUENCE 860 AA; 95455 MW; B336F868 CRC32;

Query Match 19.7%; Score 79; DB 1; Length 860;
Best Local Similarity 28.0%; Pred. No. 2.10e+00;
Matches 7; Conservative 10; Mismatches 8; Indels 0; Gaps 0;

Db 1 MSKFLTDFSNFIRVAFKYNSEAR 25
      | | | | | : | : | : | : | : |
QY 1 MEKFMAEFGQYVQTPFLSESNSVR 25

RESULT 10
ID P75688 PRELIMINARY; PRT; 70 AA.
AC P75688;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE FROM BASES 311709 TO 333910
DE (SECTION 27 OF 400) OF THE COMPLETE GENOME (SECTION 27 OF 400).
DE ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OS ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K-12;
RX MEDLINE; 97426617.

```

[illegible]

```

FT  NON_TER      1
FT  NON_TER      275
SQ  SEQUENCE      275 AA; 30495 MW; 9BE1EE8B CRC32;

Query Match      19.9%; Score 80; DB 14; Length 275;
Best Local Similarity 37.8%; Pred. No. 1.50e+00;
Matches 17; Conservative 11; Mismatches 13; Indels

Db  140 VPSPFTTELDGVLRL-RFAPPCKPLLREVSFRVGLHYPVGSQ 183
    | : | : | : | : | : | : | : | : | : | : | : | : |
Qy  13 VQTP-FLSENSVRYKISITAGSC-PLSTAGPSY-VKFDQNPVGSQ 54

RESULT 5
ID  O01993
IC  O01993

```

DT 01-JUL-1997 (TREMREL. 04, CREATED)  
DT 01-JUL-1997 (TREMREL. 04, LAST SEQUENCE UPDATE)  
DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)  
DE VC27A7L.1 PROTEIN.

GN CAENORHABDITIS ELEGANS.  
VC27A7/L.I.  
OS  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITINA; RHABDITOIDEA; RHABDITINAE. PELODERINAE. CAENORHABDITINAE.

RP SEQUENCE FROM N.A.  
RA WHITE S.;  
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFFEN L., ROOPRA A., SAUNDERS D.; SHOWNKEEN R.,  
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RL elegans.";  
RL NATURE 368:32-38(1994).  
DR EMBL; 295621; E1349927; -;  
SO SEQUENCE 357 AB; 41307 MW; DDBJ4593 C98232.

Query Match	19.9%	Score 80:	DB 5:	Length 357;
Best Local Similarity	45.88;	Pred. No.	1.50e+00;	
Matches	11:	Conservative	7:	Mismatches 6:
				Indels 0:
				Canc 0:

```

Db      211 RSYLETPFYRDYNSDVYKVPILGS 234
        ::::| | : | | | | |
Ov      10 OGCVOTPEI SEFSNVPEYFISAGS 33

```

RESULT	6	
ID	Q07034	PRELIMINARY; PRT; 802 AA.
AC	Q07034;	
DT	01-NOV-1996	(TREMREL. 01, CREATED)
DT	01-NOV-1996	(TREMREL. 01, LAST SEQUENCE UPDATE)
DT	01-JAN-1999	(TREMREL. 09, LAST ANNOTATION UPDATE)
DE	RNA BINDING PROTEIN.	
OS	SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).	
OC	EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;	
OC	SACCHAROMYCETACEAE; SACCHAROMYCES.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	SUGIMOTO K., MATSUMOTO K., KORNBURG R.D., REED S.I., WITTENBERG C.;	
RT	"HMDI of the yeast Saccharomyces cerevisiae encodes a putative	
RT	RNA/single strand DNA-binding protein, whose overexpression reduces	
RL	the expression of a Gl cyclin, CLN3."	
RL	SUBMITTED (AUG-1994) TO EMBL/GENBANK/DDBJ DATA BANKS	



\*\*\*\*\*  
M P S R C H \_ p p (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:17:15 1999; MasPar time 6.29 Seconds  
314.294 Million cell updates/sec

Tabular output not generated.

Title: >PCT-US99-13024-2  
Description: (1-93) from PCTUS9913024.pep (6 of 12)  
Perfect Score: 634  
Sequence: 1 MEKFMAEFGQGYVQTPTFLUSE.....DSKSYAFSTNSDTSAAFVS 93

Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 27.254; Variance 116.070; scale 0.235

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	300	47.3	65 19	R97371 Phage T4 ORFX gene pr
2	96	15.1	1693 35	W76368 Hepatitis E virus hol
3	92	14.5	520 1	P94617 Neutral protease enco
4	92	14.5	521 3	P51009 Sequence of neutral p
5	87	13.7	242 19	P95000 Bacillus subtilis lic
6	87	13.7	623 35	R05773 Amino acid sequence o
7	86	13.6	239 1	R06621 Hybrid (1,3-1,4)-pre-
8	86	13.6	1693 34	W71209 Protein encoded by OR
9	86	13.6	1693 19	R91813 Hepatitis E virus str
10	86	13.6	1693 10	R51264 HEV strain protein en
11	86	13.6	1693 37	W81519 Hepatitis E virus (HE
12	86	13.6	1693 36	W80196 Protein encoded by OR
13	85	13.4	633 35	W75774 Amino acid sequence o
14	84	13.2	1693 3	R14618 Protein encoded by OR
15	83	13.1	613 36	W73122 A. thaliana ethylene
16	83	13.1	613 15	R74632 QETR ethylene respons

17	82	12.9	633 35	W75775 Amino acid sequence o
18	81	12.8	345 10	R52824 GTP-cyclohydrolase II
19	81	12.8	3011 8	R40120 HCV genomic amino aci
20	81	12.8	3011 12	R66995 Hepatitis C virus gen
21	80	12.6	170 1	P90150 Sequence of hepatitis
22	80	12.6	411 16	R90934 HCV NS5 domain antige
23	80	12.6	499 38	W67010 HCV non-structural pr
24	80	12.6	509 2	R06519 Microspore-specific c
25	80	12.6	516 6	R33441 HCV NS5 encoded by pH
26	80	12.6	516 6	R33595 HCV CKS-NS5E recombin
27	80	12.6	516 7	R33633 HCV CKS-NS5E fusion p
28	80	12.6	516 4	R21566 HCV CKS-NS5E - pHCV-4
29	80	12.6	1766 1	P92041 Sequence encoded in t
30	80	12.6	1786 1	P90158 Protein sequence of h
31	80	12.6	2301 1	P92047 Sequence encoded in t
32	80	12.6	2435 5	R25135 HCV polypeptide 1
33	80	12.6	2436 5	R28582 HCV amino acid sequen
34	80	12.6	2436 1	P92050 Sequence encoded in t
35	80	12.6	2462 1	P90288 Peptide encoded by co
36	80	12.6	2772 2	R08123 Hepatitis C virus pol
37	80	12.6	2816 7	R34009 HCV-1 polypeptidein.
38	80	12.6	2894 13	R70230 Composite hepatitis C
39	80	12.6	2894 5	R24440 Composite HCV HC-J1/C
40	80	12.6	2955 2	R08124 Hepatitis C virus put
41	80	12.6	3011 4	R21519 Compiled HCV sequence
42	80	12.6	3011 6	R31621 Hepatitis C virus (HC
43	80	12.6	3011 26	W34480 HCV polypeptidein.
44	80	12.6	3011 16	R90931 Hepatitis C virus pol
45	80	12.6	3011 28	W40038 HCV polypeptidein.

ALIGNMENTS

RESULT 1  
ID R97371 standard; Protein; 65 AA.  
AC R97371;  
DE 07-JAN-1997 (first entry)  
DT Phage T4 ORFX gene product. gp34.  
KW Phage T4; tail fibre protein; nanotechnology; nano-structure;  
OS filter; molecular sieve.  
PN Bacteriophage T4.  
PN W09611947-A1.  
PD 25-APR-1996.  
PF 13-OCT-1995; U13023.  
PR 13-OCT-1994; US-322760.  
PA (GOLD/) GOLDBERG E B.  
PI Goldberg EB.  
DR WPI; 96-221942/22.  
DR N-PSDB; T29053.  
PT New proteins derived from T4 phage tail fibre proteins - that can self assemble into nano-structure(s), useful as filters etc, also corresponding DNA  
PS Claim 7; Fig 7; 83pp; English.  
CC A protein (R97371) of unspecified function is the product of open reading frame X of the tail fibre protein gene region (see also T29053) of phage T4. This gene region also includes open reading frames for tail fibre proteins (see also R97370 and R97372-74).  
CC Tail fibre proteins (native or modified) can be produced in large quantities in microbial cells and used as building blocks of strong, stable nanostructures.  
SQ Sequence 65 AA;  
Query Match 47.3%; Score 300; DB 19; Length 65;  
Best Local Similarity 82.1%; Pred. No. 1.86e-18;  
Matches 46; Conservative 3; Mismatches 7; Indels 0; Gaps 0;  
Db 1 mekfmaefggqgyvqtptfluse.....DSKSYAFSTNSDTSAAFVSgtf 56  
QY 1 MEKFMAEFGQGYVQTPTFLSESNVRYKISITAGSCPLSTAGPSYVKFDNPVGSQTF 56  
RESULT 2  
ID W76368 standard; Protein; 1693 AA.

AC W76368;  
 DT 03-DEC-1998 (first entry)  
 DE Hepatitis E virus hollow particle protein #1.  
 KW Hollow particle protein; virus; antibody; detection; immunoassay;  
 OS infection.  
 FH Hepatitis virus.  
 FT Key  
 FT Protein 1..1693  
 FT /note= "Partial sequence"  
 PN J10234383-A.  
 PD 08-SEP-1998.  
 PF 28-FEB-1997; 062445.  
 PR 28-FEB-1997; JP-062445.  
 PA (DENK-) DENKA SEIKEN KK.  
 PA (KOKU-) KOKURITSU YODO EISEI KENKYUSHO.  
 DR WPI: 98-535037/46.  
 DR N-PSDB: V61687  
 PT Hepatitis E virus hollow particle poly:peptide(s) and nucleic acids  
 PT encoding it - useful for more accurate detection of HEV in samples,  
 PT using immuno-assays and nucleic acid hybridisation  
 PS Claim 10; Page 17-24; 29pp; Japanese.  
 CC This sequence represents a Hepatitis E viral hollow particle protein.  
 CC This polypeptides can be used to raise antibodies to detect HEV  
 CC infection in samples, e.g. by immuno-assay based techniques, and the  
 CC nucleic acid can be used for the same in nucleic acid hybridisation  
 CC assays. The polypeptides and nucleic acids allow more accurate  
 CC detection of HEV than previously possible.  
 CC Sequence 1693 AA;  
 SQ  
 Query Match 15.1%; Score 96; DB 35; Length 1693;  
 Best Local Similarity 26.9%; Pred. No. 3.01e+00;  
 Matches 18; Conservative 21; Mismatches 25; Indels 3; Gaps 3;  
 Db 572 frtsvdgavleangperynisfdaqstmaagpfspyaasaglevryvaagldhrav 631  
 Qy 8 FGQGVQVFPFLESNSVRYKTSIAGCPLSTAGPSYVFQDNPVGSQT-F-SAGUHLR-V 64  
 Db 632 fapgvsp 638  
 Qy 65 FDPSTGA 71  
 RESULT 3  
 ID P94617 standard; protein: 520 AA.  
 AC P94617;  
 DT 21-JUN-1990 (first entry)  
 DE Neutral protease encoded by npr gene.  
 KW Protease; expression systems; subtilin; neutral protease; ds.  
 OS Bacillus amyloliquefaciens.  
 FH Key  
 FT Location/Qualifiers  
 FT protein 221..520  
 FT /note="Mature"  
 FT region 27..520  
 FT /note="PRO-"  
 PN US4801537-A.  
 PD 31-JAN-1989.  
 PF 29-MAR-1985; 717800.  
 PR 29-MAR-1985; US-717800.  
 PA (GENE-) Genex Corp.  
 PI Nagarajan V, Rhodes CS, Banner CDB;  
 DR WPI: 89-053639/07.  
 DR N-PSDB: N91114.  
 PT Vectors for expression of polypeptide(s) in Bacillus -  
 PT contg. promoter and regulatory regions which control expression  
 PT and secretion of protease(s) in Bacillus.  
 PS Disclosure; p: English.  
 CC Claimed replicon comprises a promoter and regulatory regions, capable of  
 CC expressing alkaline and neutral protease genes.  
 CC Sequence 520 AA;  
 Query Match 14.5%; Score 92; DB 1; Length 520;  
 Best Local Similarity 27.8%; Pred. No. 6.20e+00;  
 Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps 2;

Db 239 lssesgkylvldlskptgtqitydlqnreynlp-gtlvssttngftssgraa 291  
 Qy 36 LSTAGPSIV-KFQDNPVGSQTFSSAGLHLRVFDPSTGALVDSKSYAFSTSDTTS 88  
 RESULT 4  
 ID P51009 standard; Protein: 521 AA.  
 AC P51009;  
 DT 01-DEC-1991 (first entry)  
 DE Sequence of neutral protease encoded by the npr[BamP] gene.  
 KW Bacillus expression vector; secretion vector.  
 OS Bacillus amyloliquefaciens.  
 FH Key  
 FT Location/Qualifiers  
 FT peptide 1..27  
 FT /label= signal  
 FT peptide 28..221  
 FT /label= Pro sequence  
 FT protein 222..521  
 FT misc\_difference 1  
 FT /label= fMet  
 PN EP-133756-A.  
 PD 06-MAR-1985.  
 PF 06-JUL-1984; 304662.  
 PR 06-JUL-1983; US-511198.  
 PR 08-JUN-1984; US-618902.  
 PR 29-MAR-1985; US-717800.  
 PA (GENE-) GENEX CORP.  
 PI Nagarajan V, Banner CDB, Rhodes CS;  
 DR WPI: 85-057299/10.  
 DR N-PSDB: N50542.  
 PT Replicable plasmidic expression vector - for transformation of  
 PT Bacillus to direct expression of poly:peptide  
 PS Disclosure; Fig 4; 37pp; English.  
 CC The inventors claim a vector comprising a replicable plasmid  
 CC containing the promoter and regulatory region of a gene selected  
 CC from apr[BamP] and npr[BamP], for transformation of Bacillus to  
 CC direct expression of polypeptide.  
 CC Sequence 521 AA;  
 Query Match 14.5%; Score 92; DB 3; Length 521;  
 Best Local Similarity 27.8%; Pred. No. 6.20e+00;  
 Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps 2;

Db 240 lssesgkylvldlskptgtqitydlqnreynlp-gtlvssttngftssgraa 292  
 Qy 36 LSTAGPSIV-KFQDNPVGSQTFSSAGLHLRVFDPSTGALVDSKSYAFSTSDTTS 88  
 RESULT 5  
 ID P95000 standard; Protein: 242 AA.  
 AC P95000;  
 DT 12-FEB-1997 (first entry)  
 DE Bacillus subtilis lichenase.  
 KW Beer production; fermentation; barley; beta-glucan; hydrolysis;  
 OS lichenase.  
 OS Bacillus subtilis Y-25.  
 PN J01067181-A.  
 PD 13-MAR-1989.  
 PF 08-SEP-1987; 224615.  
 PR 08-SEP-1987; JP-224615.  
 PA (ASAK) ASAHI BREWERIES KK.  
 DR WPI: 89-119863/16.  
 DR N-PSDB: N95000.  
 PT Recombinant plasmid used in beer prodn. - obtd. by integrating  
 PT lichenase gene derived from Bacillus subtilis, into vector  
 PS Disclosure; Fig 4; 7pp; Japanese.  
 CC The lichenase gene from Bacillus subtilis Y-25 is used for  
 CC transforming Bacillus hosts so that they show increased lichenase  
 CC expression. The recombinant lichenase enzyme produced by the  
 CC transformants is useful in beer production for decomposing beta-  
 CC glucan from barley.  
 CC Sequence 242 AA;



Query Match 13.6%; Score 86; DB 34; Length 1693;  
Best Local Similarity 25.4%; Pred. No. 1.80e+01;  
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps

```
Db 572 frtsfdgavletngperhnlfsdasqstmaagpfsityaasaaglevryvaagldhrav 631
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy 8 FGQGYVQTPLSESNRYKISIAAGSCPLSTAGPSYKFDQNPVGSQT-F-SAGLHLR-V 64

Db 632 fapgvsp 638
| : : :
Qy 65 FDPSTGA 71

RESULT 9
ID R91813 standard; Protein; 1693 AA.
AC R91813;
DT 26-NOV-1996 (first entry)
DE Hepatitis E virus strain SAR-55 ORF-1.
KW Hepatitis E virus; HEV; SAR-55 strain; enteric transmission;
KW structural region; antigen: detection; antibody; vaccine;
KW immunisation; infection.
OS Hepatitis E virus.
FH Key Location/Qualifiers
FT misc_difference 1238
FT /note= "corresponding codon CAG"
FT misc_difference 1455..1693
FT /note= "10 bp nucleic acid sequence TGGTNTTYGA
FT has to be inserted between nucleotides
FT 4390..4391 of T27394 before these amino
FT acid residues can be decoded"
PN WO9610580-A2.
PD 11-APR-1996.
PF 03-OCT-1995: U13102.
PR 03-OCT-1994: US-316765.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Emerson SU, Purcell RH, Tsarev SA;
DR WPI: 96-209320/21.
DR N-PSDB: T27394.
PT Isolated and purified hepatitis E virus strain SAR-55 DNA - encodes
PT antigenic protein useful in diagnosis, prophylaxis and treatment of
PT hepatitis E virus infection
PS Disclosure: Pages 9-13: 121pp; English.
CC The present sequence is the protein prod. of ORF-1 from the
CC hepatitis E virus (HEV) strain SAR-55, which was implicated in an
CC enterically transmitted non-A, non-B hepatitis in Pakistan. The
CC protein encoded by the structural region of the virus (i.e. ORF-2),
CC which is capable of forming HEV like particles, is useful for the
CC detection of HEV antibodies (pref. IgG or IgM) in blood, plasma,
CC sera, cerebrospinal fluid, tissue, urine or pleural fluid. The
CC protein, and anti-HEV antibodies generated using the protein, can
CC also be used in vaccines for immunising an animal against HEV
CC infection. The protein is identified as a band of greater than
CC 50 kD following SDS-PAGE of cell lysates of insect cells infected
CC with a HEV ORF-2 contg. baculovirus, i.e. the claimed recombinant
CC expression vectors pPIC9-1779, -1780 and -1781.
SQ Sequence 1693 AA;

Query Match 13.6%; Score 86; DB 19; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.80e+01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 frtsfdgavletngperhnlfsdasqstmaagpfsityaasaaglevryvaagldhrav 631
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy 8 FGQGYVQTPLSESNRYKISIAAGSCPLSTAGPSYKFDQNPVGSQT-F-SAGLHLR-V 64

Db 632 fapgvsp 638
| : : :
Qy 65 FDPSTGA 71

RESULT 10
ID R51264 standard; Protein; 1693 AA.
AC R51264;
DT 21-OCT-1994 (first entry)
DE HEV strain protein encoded by ORF-1.
KW Hepatitis E virus; HEV; strain SAR-55; open reading frame; ORF;
KW antibody; detection; diagnosis; primates; stool suspension.
```

```
OS Hepatitis E virus strain SAR-55.
PN WO9406913-A.
PD 31-MAR-1994.
PF 17-SEP-1993: U08849.
PR 18-SEP-1992: US-947263.
PA (USSH ) US SEC DEPT HEALTH.
PI Emerson SU, Purcell RH, Tsarev SA;
DR WPI: 94-118462/14.
DR N-PSDB: Q45197.
PT Purified hepatitis E strain SAR-55 virus - used to develop prods.
PT for use in detection, diagnosis, vaccines and therapy of
PT hepatitis E virus infection
PS Disclosure: Page 57-62: 114pp; English.
CC The sequences given in R51264-66 are encoded by the hepatitis E virus
CC (HEV) strain SAR-55. The cDNA sequence contains three open reading
CC frames (ORFs). These proteins can be used to stimulate the production
CC of protective antibodies upon injection into a mammal that would serve
CC to protect the mammal upon challenge with wild type HEV. The proteins
CC can be used for detection and diagnosis of HEV infection. The HEV
CC SAR-55 cDNA was isolated from primates inoculated with stool
CC suspensions obtained from hepatitis E patients.
SQ Sequence 1693 AA;

Query Match 13.6%; Score 86; DB 10; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.80e+01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 frtsfdgavletngperhnlfsdasqstmaagpfsityaasaaglevryvaagldhrav 631
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy 8 FGQGYVQTPLSESNRYKISIAAGSCPLSTAGPSYKFDQNPVGSQT-F-SAGLHLR-V 64

Db 632 fapgvsp 638
| : : :
Qy 65 FDPSTGA 71

RESULT 11
ID W81519 standard; Protein; 1693 AA.
AC W81519;
DT 02-FEB-1999 (first entry)
DE Hepatitis E virus (HEV) ORF-1 protein.
KW Hepatitis E virus; HEV; SAR-55; diagnostic agent; vaccine; antibody;
KW passive immunisation; open reading frame; ORF.
OS Hepatitis E virus.
FH Key Location/Qualifiers
FT Misc_difference 1238
FT /note= "encoded by CAG"
FT Misc_difference 1244
FT /note= "encoded by CAG"
FT Misc_difference 1352
FT /note= "encoded by GTG"
FT Misc_difference 1662
FT /note= "encoded by GGC"
PN WO9846761-A1.
PD 22-OCT-1998.
PR 09-APR-1998; U07418.
PR 11-APR-1997: US-840316.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Emerson SU, Purcell RH, Robinson RA, Tsarev SA;
DR WPI: 98-568733/48.
DR N-PSDB: V71604.
PT New hepatitis E virus DNA from Pakistani strain SAR-55 - used for,
PT e.g. developing products for diagnosis of, and vaccination against
PT hepatitis E virus infection
PS Disclosure: Pages 11-15: 204pp; English.
CC This represents a hepatitis E virus (HEV) open reading frame (ORF)-1
CC protein encoded by a DNA sequence designated SAR-55. SAR-55 also encodes
CC HEV ORF-2 and ORF-3 proteins. A host organism transformed or transfected
CC with a recombinant expression vector containing the SAR-55 nucleic acid
CC can be used to produce the HEV proteins, especially ORF-2 protein. The
CC recombinant HEV proteins can be used as diagnostic agents and as vaccines
CC for use against HEV infection. The detection of antibodies specific for
CC HEV can be used for the diagnosis of infection and diseases caused by
```





\*\*\*\*\*

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

MPsrch_pp      protein - protein database search, using Smith-Waterman algorithm
Run on:        Wed Sep 1 16:31:44 1999;  MasPar time 5.61 Seconds
               126.558 Million cell updates/sec

```

120.538 MILLION cell updates/sec

Tabular output not generated.

```
>PCT-US99-13024-2
Description: (81-93) from PCTUS9913024.pep (12 of 12)
Perfect Score: 68
Sequence: 1 STSNDTTSAAEVS 13
```

Sequence: 1 STSNPTTSAAAFVS 13

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: sptrembl9

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 21.936; Variance 22.856; scale 0.960

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	54	79.4	101	11	088517	CALCIUM CHANNEL BETA 1	7.52e-02
2	54	79.4	579	4	015331	L-TYPE CALCIUM CHANNEL	7.52e-02
3	47	69.1	241	5	020870	SIMILAR TO GLUTAMATE R	4.20e+00
4	47	69.1	982	5	Q09532	HYPOTHETICAL 110.9 KD	4.20e+00
5	47	69.1	2100	3	P87112	HYPOTHETICAL 110.9 KD	4.20e+00
6	46	67.6	268	5	Q94484	PURF TRANSCRIPTION	7.24e+00
7	45	66.2	337	5	Q19474	OUTF DG1007 (FRAGMENT)	7.24e+00
8	45	66.2	583	3	Q12372	F15A2.4 PROTEIN.	1.24e+01
9	44	64.7	37	5	Q26818	CHROMOSOME XII READING	1.24e+01
10	44	64.7	53	5	Q3582	SURFACE ANTIGEN (FRAG	2.09e+01
11	44	64.7	236	11	Q00900	SIMILAR TO ADENOSINE A	2.09e+01
12	44	64.7	371	10	085584	POTASSIUM CHANNEL PROT	2.09e+01
13	44	64.7	461	2	Q68923	HYPOTHETICAL 40.8 KD P	2.09e+01
14	44	64.7	488	5	Q43137	GLYCERALDEHYDE-3-PHOS	2.09e+01
15	44	64.7	545	3	P87306	REVERSE TRANSCRIPTASE	2.09e+01
16	44	64.7	575	5	Q26040	HYPOTHETICAL 62.5 KD P	2.09e+01
17	44	64.7	611	11	Q32886	SHAL 1 POTASSIUM CHANN	2.09e+01
18	44	64.7	635	5	Q36922	VOLTAGE-GATED K+ CHANN	2.09e+01
19	44	64.7	636	11	W03F9.4	W03F9.4 PROTEIN.	2.09e+01
20	44	64.7	636	4	P070622	KV4.3 POTASSIUM CHANNE	2.09e+01
21	44	64.7	636	4	O60576	KV4.3 POTASSIUM CHANNE	2.09e+01

RP SEQUENCE FROM N.A.  
RA POWERS P.A., GREGG R.G., HOGAN K.;  
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U86953; G2155256; JOINED.  
DR EMBL: U86954; G2155256; JOINED.  
DR EMBL: U86955; G2155256; JOINED.  
DR EMBL: U86961; G2155256; -.  
DR EMBL: U86962; G2155256; -.  
DR EMBL: U86963; G2155256; JOINED.  
DR EMBL: U86956; G2155256; JOINED.  
DR EMBL: U86959; G2155256; JOINED.  
DR EMBL: U86960; G2155256; JOINED.  
DR EMBL: U86957; G2155256; JOINED.  
DR EMBL: U86958; G2155256; JOINED.  
DR PFAM: PF00018; SH3; 1.  
DR PFAM: PF00774; Ca\_channel\_B; 1.  
SQ SEQUENCE 579 AA; 63705 MW; 651614EF CRC32;  
  
Query Match 79.4%; Score 54; DB 4; Length 579;  
Best Local Similarity 75.0%; Pred. No. 7.52e-02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
Db 44 STSDTTSNSFV 55  
:|||||:|  
Qy 81 STSNDTTSAAFV 92  
  
RESULT 3  
ID Q20870; PRELIMINARY; PRT; 241 AA.  
AC Q20870;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE SIMILAR TO GLUTAMATE RECEPTORS.  
GN F56D1.7.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA;  
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94150718.  
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON A., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "2.2 kb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL NATURE 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA CHISOOE S., WILSON R.;  
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA WATERSTON R.;  
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U39997; G1055122; -.  
SQ SEQUENCE 241 AA; 27445 MW; F0977197 CRC32;  
  
Query Match 69.1%; Score 47; DB 5; Length 241;  
Best Local Similarity 80.0%; Pred. No. 4.20e+00;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Db 59 GNDTISAFAV 68  
:|||||  
Qy 83 SNDTTSAAFV 92

RESULT 6  
ID Q94484; PRELIMINARY; PRT; 268 AA.  
AC Q94484;  
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
DT 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)  
DE ORF DG1007 (FRAGMENT).  
OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).  
  
Query Match 69.1%; Score 47; DB 3; Length 2100;  
Best Local Similarity 66.7%; Pred. No. 4.20e+00;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
Db 1601 SATNDETKAAAFV 1612  
:|||||  
Qy 81 STSNDTTSAAFV 92

RESULT 4  
ID Q09532; PRELIMINARY; PRT; 982 AA.  
AC Q09532;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 110.9 KD PROTEIN F07F6.6 IN CHROMOSOME II.  
GN F07F6.6.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA;  
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL NZ;  
RA CHISOOE S.;  
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- SIMILARITY: TO GLUTAMATE RECEPTOR.  
DR EMBL: U23486; G746452; -  
DR WORMPEP; F07F6.6; CE01898.  
DR PFAM: PF00060; lig\_chan; 1.  
KW HYPOTHETICAL PROTEIN.  
FT DOMAIN 827 830 POLY-ARG.  
SQ SEQUENCE 982 AA; 110892 MW; B4285770 CRC32;  
  
Query Match 69.1%; Score 47; DB 5; Length 982;  
Best Local Similarity 80.0%; Pred. No. 4.20e+00;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Db 59 GNDTISAFAV 68  
:|||||  
Qy 83 SNDTTSAAFV 92  
  
RESULT 5  
ID P87112; PRELIMINARY; PRT; 2100 AA.  
AC P87112;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE PUTATIVE TRANSCRIPTIONAL REGULATOR C20G8.06.  
GN SPAC20G8.06.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;  
OC SCHIZOSACCHAROMYCETES; SCHIZOSACCHAROMYCETACEAE;  
OC SCHIZOSACCHAROMYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972;  
RA BADCOCK K., CHURCHER C.M., WOOD V., BARRELL B.G., RAJANDREAM M.A.;  
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
CC -1- SIMILARITY: TO YEAST NOT1.  
DR EMBL: Z95334; E315491; -  
KW HYPOTHETICAL PROTEIN; NUCLEAR PROTEIN; TRANSCRIPTION REGULATION.  
SQ SEQUENCE 2100 AA; 237132 MW; 9A710DB4 CRC32;  
  
Query Match 69.1%; Score 47; DB 3; Length 2100;  
Best Local Similarity 66.7%; Pred. No. 4.20e+00;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
Db 1601 SATNDETKAAAFV 1612  
:|||||  
Qy 81 STSNDTTSAAFV 92

OC EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AX4;  
RA LOOMIS W.F.;  
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U56911; G1519534;  
FT NON\_LTER 268 268  
SQ SEQUENCE 268 AA; 30503 MW; 04E723F5 CRC32;  
  
Query Match 67.6%; Score 46; DB 5; Length 268;  
Best Local Similarity 53.8%; Pred. No. 7.24e+00;  
Matches 4; Mismatches 2; Indels 0; Gaps 0;  
  
Db 15 TTNNTTPNKEVS 27  
:|:|:|:|  
QY 81 STSNDTTSAAVFS 93  
  
RESULT 7  
ID Q19474 PRELIMINARY; PRT; 337 AA.  
AC Q19474;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)  
DE F15A2.4 PROTEIN.  
GN F15A2.4.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA;  
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA GREGORY J.;  
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SHALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans";  
RL NATURE 368:32-38(1994).  
DR EMBL: 270207; E1345607;  
SQ SEQUENCE 337 AA; 38483 MW; ABD004B8 CRC32;  
  
Query Match 66.2%; Score 45; DB 5; Length 337;  
Best Local Similarity 66.7%; Pred. No. 1.24e+01;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
Db 317 SKHNTTSATFV 328  
| | | | | | | |  
QY 81 STSNDTTSAAVFS 92  
  
RESULT 8  
ID Q12372 PRELIMINARY; PRT; 583 AA.  
AC Q12372;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE CHROMOSOME XII READING FRAME ORF YLL061W.  
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;  
OC SACCHAROMYCETACEAE; SACCHAROMYCES.  
RN [1]

RP SEQUENCE FROM N.A.  
RA WEDLER H., WEDLER E., SCHARFE M., WAMBUIT R.;  
RN SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA MIPS;  
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C;  
RA WEDLER H., WAMBUIT R.;  
RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: Z73166; E245763;  
DR EMBL: Z47973; G642323;  
DR PFAM: PF00324; aa\_permeases; 1.  
SQ SEQUENCE 583 AA; 64217 MW; BDE5153C CRC32;  
  
Query Match 66.2%; Score 45; DB 3; Length 583;  
Best Local Similarity 50.0%; Pred. No. 1.24e+01;  
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
  
Db 330 SSEDITASPFV 341  
|:|:| |::| |  
QY 81 STSNDTTSAAVFS 92  
  
RESULT 9  
ID Q26818 PRELIMINARY; PRT; 37 AA.  
AC Q26818;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE SURFACE ANTIGEN (FRAGMENT).  
OS TRYPAPOSOMA BRUCEI BRUCEI.  
OC EUKARYOTA; EUGLENOZOA; KINETOPLASTIDA; TRYPAPOSOMATIDAE; TRYPAPOSOMA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GAMBIENSE;  
RX MEDLINE: 83169719.  
RA MERRITT S.C., TSCHUDI C., KONIGSBERG W.H., RICHARDS F.F.;  
RT "Reverse transcription of trypanosome variable antigen mRNAs  
initiated by a specific oligonucleotide primer";  
RL PROC. NATL. ACAD. SCI. U.S.A. 80:1536-1540(1983).  
DR EMBL: V01391; G10560;  
FT NON\_TER 1 1  
SQ SEQUENCE 37 AA; 4068 MW; 551442D4 CRC32;  
  
Query Match 64.7%; Score 44; DB 5; Length 37;  
Best Local Similarity 63.6%; Pred. No. 2.09e+01;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
Db 14 TSNTTASNSFV 24  
| | | | | | | |  
QY 82 TSNDTTSAAVFS 92  
  
RESULT 10  
ID Q23582 PRELIMINARY; PRT; 53 AA.  
AC Q23582;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE SIMILAR TO ADENOSINE A3 RECEPTOR. NCBI GI: 1072232.  
GN ZK721.5.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA;  
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,

RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOHLDMAN P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans.";  
 RL NATURE 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MILLER N.;  
 RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA WATERSTON R.;  
 RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: U40951; G107232;  
 SQ SEQUENCE 53 AA; 5833 MW; C4DA122E CRC32;

Query Match 64.7%; Score 44; DB 5; Length 53;  
 Best Local Similarity 58.3%; Pred. No. 2.09e+01;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 6 STSNSTRVSFV 17  
 ||||| : :||  
 Oy 81 STSNDTSAFVS 92

RESULT 11  
 ID Q00090 PRELIMINARY; PRT; 236 AA.  
 DT 01-NOV-1996 (TREMREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
 DE POTASSIUM CHANNEL PROTEIN (FRAGMENT).  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA RUDY B., KENTROS C., DE MIERA E.;  
 RL MOL. CELL. NEUROSCI. 2:89-102(1991).  
 CC -!- FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT POTASSIUM  
 ION PERMEABILITY OF EXCITABLE MEMBRANES.  
 CC ASSUMING OPENED OR CLOSED CONFORMATIONS IN RESPONSE TO THE  
 VOLTAGE DIFFERENCE ACROSS THE MEMBRANE, THE PROTEIN FORMS A  
 POTASSIUM-SELECTIVE CHANNEL THROUGH WHICH K+ IONS MAY PASS IN  
 ACCORDANCE WITH THEIR ELECTROCHEMICAL GRADIENT.  
 CC MAY PLAY A ROLE IN THE NERVOUS SYSTEM AND IN THE REGULATION OF  
 BEATING FREQUENCY IN PACEMAKER CELLS.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -!- THE SEGMENT S4 IS PROBABLY THE VOLTAGE-SENSOR AND IS  
 CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT  
 EVERY THIRD POSITION.  
 CC -!- THE AMINO TERMINUS MAY BE IMPORTANT IN DETERMINING THE RATE OF  
 INACTIVATION OF THE CHANNEL WHILE THE TAIL MAY PLAY A ROLE IN  
 MODULATION OF CHANNEL ACTIVITY AND/OR TARGETING OF THE CHANNEL TO  
 SPECIFIC SUBCELLULAR COMPARTMENTS.  
 CC -!- SIMILARITY: THIS CHANNEL PROTEIN BELONGS TO THE A-TYPE POTASSIUM  
 CURRENT CLASS.  
 CC -!- SIMILARITY: BELONGING TO THE SHIV GENE SUBFAMILY, A HOMOLOGUE OF  
 THE SHAKER GENE SHAL OF DROSOPHILA MELANOGASTER.  
 DR EMBL: M74898; G205041;  
 DR PFAM: PF00520; ion.trans. 1.  
 KW IONIC CHANNEL; TRANSMEMBRANE; ION TRANSPORT; VOLTAGE-GATED CHANNEL;  
 KW MULTIGENE FAMILY.  
 FT TRANSMEM 15 33 SEGMENT S1 (BY SIMILARITY).  
 FT TRANSMEM 57 78 SEGMENT S2 (BY SIMILARITY).  
 FT TRANSMEM 89 110 SEGMENT S3 (BY SIMILARITY).  
 FT TRANSMEM 118 136 SEGMENT S4 (BY SIMILARITY).

FT TRANSMEM 152 173 SEGMENT S5 (BY SIMILARITY).  
 FT TRANSMEM 213 234 SEGMENT S6 (BY SIMILARITY).  
 FT NON\_TER 236  
 SQ SEQUENCE 236 AA; 26261 MW; C4A458FB CRC32;

Query Match 64.7%; Score 44; DB 11; Length 236;  
 Best Local Similarity 50.0%; Pred. No. 2.09e+01;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 108 TNNEDVSGAFVT 119  
 |::| :|||  
 Oy 82 TSNDTSAFVS 93

RESULT 12  
 ID O65584 PRELIMINARY; PRT; 371 AA.  
 AC O65584;  
 DT 01-AUG-1998 (TREMREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 40.8 KD PROTEIN.  
 GN M359.70.  
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; TRACHEOPHYTA;  
 OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;  
 OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BEVAN M., VANDENBOL M., JALLET C., PORTELETTE D., HOHEISEL J.,  
 RA MENES H.W., MAYER K., SCHUELLER C.;  
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU ARABIDOPSIS SEQUENCING PROJECT;  
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AL022223; E1283943;  
 KW HYPOTHETICAL PROTEIN  
 SQ SEQUENCE 371 AA; 40833 MW; 5B0F5AA1 CRC32;

Query Match 64.7%; Score 44; DB 10; Length 371;  
 Best Local Similarity 38.5%; Pred. No. 2.09e+01;  
 Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Db 224 SSSGSESSSFVS 236  
 |:|::: ||||  
 Oy 81 STSNDTSAFVS 93

RESULT 13  
 ID O68923 PRELIMINARY; PRT; 461 AA.  
 AC O68923;  
 DT 01-AUG-1998 (TREMREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
 DE GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE HOMOLOG.  
 GN GAPX.  
 OS STREPTOMYCES ROSEOFULVUS.  
 OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;  
 OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94237486.  
 RA BIBB M.J., SHERMAN D.H., OMURA S., HOPWOOD D.A.;  
 RT "Cloning, sequencing and deduced functions of a cluster of  
 RT streptomycetes genes probably encoding biosynthesis of the polyketide  
 RT antibiotic frenolicin.";  
 RL GENE 142:31-39(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA REEVES C.D., SOLIDAY C.L.;  
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AF058302; G3170587;  
 SQ SEQUENCE 461 AA; 49966 MW; 26EC2672 CRC32;



Query Match 64.7%; Score 44; DB 2; Length 461;  
Best Local Similarity 46.2%; Pred. No. 2.09e+01;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
DB 394 TTSADAVSDFVC 406  
QY 81 STSNDTTSAAFS 93

RESULT 14  
ID 04317 PRELIMINARY; PRT; 488 AA.  
AC 04317;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE REVERSE TRANSCRIPTASE (FRAGMENT).  
GN R1 ORF.  
OS DUGESIELLA SP.  
OC EUKARYOTA; METAZOA; ARTHROPODA; CHELICERATA; ARACHNIDA; ARANEAE;  
OC MYGALOMORPHAE; THERAPHOSIDAE; DUGESIELLA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BURKE W.D., MALIK H.S., EICKBUSH T.H.;  
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AF015813; G2736044; -.  
KW RNA-DIRECTED DNA POLYMERASE.  
FT NON\_TER 1  
SQ SEQUENCE 488 AA; 55092 MW; FA67B24C CRC32;

Query Match 64.7%; Score 44; DB 5; Length 488;  
Best Local Similarity 58.3%; Pred. No. 2.09e+01;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
DB 251 SKTSDGTGAFFV 262  
QY 81 STSNDTTSAAFS 92

RESULT 15  
ID P87306 PRELIMINARY; PRT; 545 AA.  
AC P87306;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 62.5 KD PROTEIN C31F10.04C IN CHROMOSOME II.  
GN SPBC31F10.04C.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;  
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972;  
RA WOOD V., RAJANDREAM M.A., BARRELL B.G., POHL T.;  
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; Z97204; E325252; -.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 545 AA; 62479 MW; 2F16D503 CRC32;

Query Match 64.7%; Score 44; DB 3; Length 545;  
Best Local Similarity 54.5%; Pred. No. 2.09e+01;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
DB 364 SNEDTSTNEVT 374  
QY 83 SNDTTSAAFS 93

Search completed: Wed Sep 1 16:32:18 1999  
Job time : 34 secs.

```

SQ SEQUENCE 1036 AA; 109310 MW; CCDE5BE8 CRC32;
Query Match 70.6%; Score 48; DB 1; Length 1036;
Best Local Similarity 53.8%; Pred. No. 1.11e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 87 SSSNNNTSNGFVA 99
    :||: || :||:
QY 81 STSNDTTSAAVFS 93

RESULT 11
ID CPPLDROAC STANDARD; PRT; 532 AA.
AC 04220;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DE CYTOCHROME P450 12B1 PRECURSOR (EC 1.14.-.-).
GN CYP12B1
OS DROSOPHILA ACANTHOPTERA (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A584.2;
RX MEDLINE; 97449781.
RA DANIELSON P.B., FOGLEMAN J.C.;
RT "Isolation and sequence analysis of cytochrome P450 12B1: the first
RT mitochondrial insect P450 with homology to 1 alpha,25 dihydroxy-D3
RT 24-hydroxylase."
RL INSECT BIOCHEM. MOL. BIOL. 27:595-604(1997).
CC -!- FUNCTION: PROBABLY INVOLVED IN STEROID HORMONES BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U78485; G2674278; -
CC DR FLYBASE; FBgn0022933; Dacc\Cyp12b1.
CC DR PROSITE; PS00086; CYTOCHROME P450; 1.
CC KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; HEME;
CC MW MITOCHONDRION; TRANSIT PEPTIDE.
CC FT TRANSIT 1 ? MITOCHONDRION (BY SIMILARITY).
CC FT CHAIN ? 532 CYTOCHROME P450 12B1.
CC FT BINDING 480 480 HEME (POTENTIAL).
CC SQ SEQUENCE 532 AA; 62031 MW; 53F97355 CRC32;

Query Match 67.6%; Score 46; DB 1; Length 532;
Best Local Similarity 77.8%; Pred. No. 3.44e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 335 DTSSAFVT 343
    ||||:|
QY 85 DTTSAAVFS 93

RESULT 12
ID YEM3_YEAST STANDARD; PRT; 1076 AA.
AC P40021;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DE CYTOCHROME P450 12B1 PRECURSOR (EC 1.14.-.-).
GN YER033C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;

```

```

CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; U67505; G1591267; -.
DR TIGR; MJ0561; -.
DR PROSITE; PS01266; ADENYLOSUCIN_SYN_1; 1.
DR PROSITE; PS00513; ADENYLOSUCIN_SYN_2; FALSE_NEG.
DR PFAM; PF00709; Adenylsucc_synth; 1.
DR HSP; P12283; IADI.
KW PURINE BIOSYNTHESIS; 24
FT NP_BIND 18 37820 MW; 8289EB89 CRC32;
SQ SEQUENCE 345 AA; 47416 MW; 8289EB89 CRC32;

Query Match 66.2%; Score 45; DB 1; Length 345;
Best Local Similarity 50.0%; Pred. No. 5.97e+00;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 200 TSKDTTASSFAA 211
QY 82 TSDNTTSAAFVS 93
||||:||||:|

RESULT 14
ID IR12_HCMVA STANDARD; PRT; 416 AA.
AC P16810;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DE 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN IRL12 (TRL12).
OS HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC BETAHERPESVIRINAE; CYTOMEGALOVIRUS.
CC [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 90269039.
RA CHEE M.S., BANKIER A.T., BECK S., BOHNI R., BROWN C.M., CERNY R.,
RA HORSNELL T., HUTCHISON C.A. III, KOUZARIDES T., MARTIGNETTI J.A.,
RA PREDDIE E., SACHWELL S.C., TOMLINSON P., WESTON K.M., BARRELL B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169."
RL CURR. TOP. MICROBIOL. IMMUNOL. 154:125-169(1990).
CC -1- SIMILARITY: BELONGS TO RL11 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; X17403; E298604; -.
DR EMBL; X17403; E298631; -.
DR PIR; S09761; S09761.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 416 AA; 47416 MW; 8289EB89 CRC32;

Query Match 66.2%; Score 45; DB 1; Length 416;
Best Local Similarity 50.0%; Pred. No. 5.97e+00;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 42 TSPNTASTTFTV 53
QY 82 TSDNTTSAAFVS 93
||||:||||:|

RESULT 15
ID POLG_BCMVN STANDARD; PRT; 3066 AA.
AC O65329;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DE 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER
DE COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KD PROTEIN

```

```

DE 1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KD PROTEIN 2 (6K2);
DE GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)
DE (EC 3.4.22.-) (49 KD PROTEINASE) (49 KD-PRO); NUCLEAR INCLUSION
DE PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);
DE COAT PROTEIN (CP)].
OS BEAN COMMON MOSAIC VIRUS (STRAIN NL-3 / MICHIGAN) (BCMV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC POTYVIRUS.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 96191623.
RA FANG G.W., ALLISON R.F., ZAMBOLIM E.M., MAXWELL D.P., GILBERTSON R.L.;
RT "The complete nucleotide sequence and genome organization of bean
RT common mosaic virus (NL3 strain).";
RL VIRUS RES. 39:13-23(1995).
CC -1- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID
CC TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
CC -1- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
CC MAY BE INVOLVED IN REPLICATION.
CC -1- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
CC -1- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.
CC -1- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
CC POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
CC INDIVIDUAL PROTEINS.
CC -1- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
CC -1- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
CC -1- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; U19287; G1373002; -.
DR PFAM; PF00271; helicase_C; 1.
DR PFAM; PF00580; RNA_dep_RNA_pol; 1.
DR PFAM; PF00767; Poty_coat; 1.
DR PFAM; PF00851; Peptidase_C6; 1.
DR PFAM; PF00863; Peptidase_C4; 1.
KW HYDROLASE; TRANSFERASE; THIOL PROTEASE; RNA-DIRECTED RNA POLYMERASE;
KW COAT PROTEIN; POLYPROTEIN; COVALENT PROTEIN-RNA LINKAGE; HELICASE;
KW ATP-BINDING.
FT CHAIN 1 ? N-TERMINAL PROTEIN.
FT CHAIN ? ? HELPER COMPONENT PROTEINASE.
FT CHAIN ? ? PROTEIN P3.
FT CHAIN ? ? 6 KD PROTEIN 1.
FT CHAIN ? ? CYTOPLASMIC INCLUSION PROTEIN.
FT CHAIN ? ? 6 KD PROTEIN 2.
FT CHAIN ? ? GENOME-LINKED PROTEIN.
FT CHAIN ? ? NUCLEAR INCLUSION PROTEIN A.
FT CHAIN ? ? NUCLEAR INCLUSION PROTEIN B.
FT CHAIN ? ? COAT PROTEIN.
FT NP_BIND 1258 1265 ATP (POTENTIAL)
SQ SEQUENCE 3066 AA; 350387 MW; 67B875D6 CRC32;

Query Match 66.2%; Score 45; DB 1; Length 3066;
Best Local Similarity 54.5%; Pred. No. 5.97e+00;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 1194 STSKESTSTSF 1204
QY 81 STSNDTTSAAF 91
||||:||||:|

Search completed: Wed Sep 1 16:31:25 1999
Job time : 11 secs.

```

(TIN)

```

Release 3.1A John F. Collins, Biocomputing Research Unit:
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp    protein - protein database search, using Smith-Waterman algorithm
Run on:      Wed Sep 1 16:13:05 1999; MasPar time 9.45 Seconds
Tabular output not generated.      450.572 Million cell updates/sec

Title:       >PCT-US99-13024-2
Description: (1-78) from PCTUS9913024.pep (5 of 12)
Perfect Score: 551
Sequence:    1 MEKFMAEFGQYVOTPFLE.....GLHRLRVDFDPSTGALVDSKSY 78

```

```

Scoring table:      PAM 150
                   Gap 11

Searched:          179066 seqs, 54579741 residues

Post-processing:    Minimum Match 0%
                   Listing first 45 summaries

Database:
  1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
  5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
  9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
  13:sp-vertebrate 14:sp-virus

Statistics:         Mean 37.051;  Variance 67.149;  scale 0.552

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query		Length	DB	ID	Description	Pred. No.
		Match	%					
1	100	18.1	1693	14	Q81876		COMPLETE GENOME SEQUEN	4.78e+03
2	96	17.4	114	14	Q41088		A065L PROTEIN.	2.00e-02
3	92	16.7	299	14	Q65574		HYPOTHETICAL 32.4 KD P	8.15e-02
4	92	16.7	300	14	Q65581		UL7 POLYPEPTIDE.	8.15e-02
5	89	16.2	566	2	Q84354		HYPOTHETICAL 63.5 KD P	2.28e-01
6	86	15.6	1593	14	Q81862		ORF 1.	6.27e-01
7	86	15.6	1693	14	Q81344		HEPATITIS E VIRUS COM	6.27e-01
8	86	15.6	1693	14	Q83444		UNNAMED PROTEIN PRODUC	6.27e-01
9	86	15.6	1693	14	Q69410		METHYL TRANSFERASE.	6.27e-01
10	84	15.2	622	2	Q45743		CRYIIC DELTA-ENDOTOXIN	1.21e+00
11	84	15.2	921	5	P90770		C34B7.2 PROTEIN.	1.21e+00
12	83	15.1	180	10	Q49984		PUPATIVE ETHYLENE RECE	1.68e+00
13	83	15.1	281	9	Q38022		ORF 1.	1.68e+00
14	83	15.1	1377	2	Q58178		GLUCOSYLTRANSFERASE.	1.68e+00
15	83	15.1	1693	14	Q39221		NONSTRUCTURAL POLYPROT	1.68e+00
16	81	14.7	196	14	Q65847		PUPATIVE COAT PROTEIN.	3.21e+00
17	81	14.7	216	2	Q67609		FLAGELLAR L-RING PROTE	3.21e+00
18	81	14.7	633	2	Q52764		INSECTICIDAL CRYSTAL P	3.21e+00
19	81	14.7	642	14	Q65848		PUPATIVE READTHROUGH P	3.21e+00
20	81	14.7	3011	14	Q39463		GENOME POLYPROTEIN.	3.21e+00

21	80	14.5	275	14	Q68467	POLYPROTEIN (FRAGMENT)	4.41e+000
22	80	14.5	275	14	Q68468	POLYPROTEIN (FRAGMENT)	4.41e+000
23	80	14.5	275	14	Q68469	POLYPROTEIN (FRAGMENT)	4.41e+000
24	80	14.5	357	5	Q01993	VC27A7L-1 PROTEIN.	4.41e+000
25	80	14.5	802	3	Q07034	RNA BINDING PROTEIN.	4.41e+000
26	80	14.5	2436	14	Q81756	POLYPROTEIN (FRAGMENT)	4.41e+000
27	79	14.3	370	3	Q04083	D9461.24P.	6.06e+000
28	79	14.3	542	5	Q17456	SIMILAR TO GLUTAMATE D	6.06e+000
29	78	14.3	860	1	Q59003	860AA LONG HYPOTHETICA	6.06e+000
30	78	14.2	70	2	P75688	FROM BASES 311709 TO 3	8.28e+000
31	78	14.2	613	10	Q38846	ETHYLENE RESPONSE SENS	8.28e+000
32	78	14.2	3011	14	Q36609	POLYPROTEIN.	8.28e+000
33	78	14.2	3011	14	Q36608	POLYPROTEIN.	8.28e+000
34	78	14.2	3011	14	Q36610	POLYPROTEIN.	8.28e+000
35	78	14.2	3011	14	Q36579	POLYPROTEIN.	8.28e+000
36	77	14.0	418	2	Q56631	LECITHINASE.	1.13e+01
37	77	14.0	470	2	Q87325	LECITHINASE.	1.13e+01
38	77	14.0	510	10	Q65322	PUTATIVE MONOSACCHARID	1.13e+01
39	77	14.0	967	2	Q34123	PEPP.	1.13e+01
40	77	14.0	1361	2	Q86617	HYPOTHETICAL 145.0 KD	1.13e+01
41	77	14.0	1464	11	Q08948	N-METHYL-D-ASPARTATE R	1.13e+01
42	77	14.0	1464	11	Q63728	N-METHYL-D-ASPARTATE R	1.13e+01
43	77	14.0	1464	4	Q12879	N-METHYL-D-ASPARTATE R	1.13e+01
44	77	14.0	2219	5	Q23388	UK1067.2 PROTEIN.	1.13e+01
45	76	13.8	821	2	Q31335	OUTER MEMBRANE PROTEIN	1.54e+01

## ALIGNMENTS

```

RESULT      1
ID          Q81876      PRELIMINARY;      PRT; 1693 AA.
AC          Q81876;
DT          01-NOV-1996 (TREMBLREL. 01, CREATED)
DT          01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT          01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE          COMPLETE GENOME SEQUENCE.
DE          HEPATITIS E VIRUS (HEV).
OS          CALICIVIRIDAE;
OC          VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE;
OC          CALICIVIRUS.
RN          [1]
RP          SEQUENCE FROM N.A.
RC          STRAIN-UIGH179;
RA          UCHIDA T.;
RL          SUBMITTED, (MAY-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
DR          EMBL: D11093; D1002342; ...
SQ          SEQUENCE 1693 AA; 184840 MW; 8952DD38 CRC32;

```

Query Match 18.1%; Score 100; DB 14; Length 1693;  
Best Local Similarity 26.9%; Pred. No. 4.78e-03;  
Matches 18; Conservative 22; Mismatches 24; Indels 3; Gaps 3

```

572 FRFSYDGAIVLEANGPERYNYLSEFSDASOSTWAAGPFSITYAASAAGLEVYRYVAAGLDHRV 631
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 8 FGQYVQTPFLESNSVRYKISIAAGSCPLSTAGPFSYVKFQDNPVGSQT-F-SAGLHLR-V 64

632 FAPGVSP 638
QY 65 FDPSTGA 71

```

RESULT	2	PRELIMINARY;	PRT;	114 AA.
ID	O41088			
AC	O41088;			
DT	01-JAN-1998 (TREMBREL 05, CREATED)			
DT	01-JAN-1998 (TREMBREL. 05, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)			
DE	A606L PROTEIN.			
GN	A606L.			
OS	PARAMECTUM BURSARIA CHLORELLA VIRUS 1 (PCV-1).			
OC	VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PHYCODNAVIRIDAE; PHYCODNAVIRUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 98022962.			

```

RA LI Y., LU Z., SUN L., ROPP S., KUTISH G.F., ROCK D.L.,
RA VAN ETEN J.L.;
RT "Analysis of 74 Kb of DNA located at the right end of the 330-kb
RT Chlorella virus PBCV-1 genome.";
RL VIROLOGY 237:360-377(1997).
DR EMBL: U42580; G2447164; -.
SQ SEQUENCE 114 AA; 13257 MW; 1B75BE43 CRC32;

Query Match 17.4%; Score 96; DB 14; Length 114;
Best Local Similarity 27.4%; Pred. No. 2,00e-02;
Matches 17; Conservative 18; Mismatches 25; Indels 2; Gaps 1;

Db 5 F5SM5N5IFRVL7ASCAL--LNSYSGRKRNRNSLNFTWSLSLMVFVSIKPSMSIK 62
QY 17 FLSENSVRYKISIAAGSCPLSTAGPSYKVFQDNVPGSQTSFAGLHRLVDFPSTGALVDSK 76
Db 63 KY 64
QY 77 SY 78

RESULT 3
ID Q65574 PRELIMINARY; PRT; 299 AA.
AC Q65574;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 32.4 KD PROTEIN.
GN UL7.
OS BOVINE HERPESVIRUS TYPE 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.
RN
RP SEQUENCE FROM N.A.
RC STRAIN=COOPER, AND JURA;
RX MEDLINE: 95313343.
RA VLCEK C., BENES V., LU Z., KUTISH G.F., PACES V., ROCK D.,
RA LETCHWORTH G.J., SCHWYZER M.;
RT "Nucleotide sequence analysis of a 30-kb region of the bovine
RT herpesvirus 1 genome which exhibits a colinear gene arrangement with
RT the UL12 to UL4 genes of herpes simplex virus.";
RL VIROLOGY 210:100-108(1995).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=JURA;
RA SCHWYZER M.;
RA LOWERY D.E., SIMARD C., BELLO L.J., THIRY E., VLCEK C.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN
RP SEQUENCE FROM N.A.
RC STRAIN=JURA;
RA SCHWYZER M.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: Z48053; G971327; -.
DR EMBL: AJ004801; E1187337; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 299 AA; 32379 MW; A3897D4A CRC32;

Query Match 16.7%; Score 92; DB 14; Length 299;
Best Local Similarity 47.2%; Pred. No. 8.15e-02;
Matches 17; Conservative 9; Mismatches 6; Indels 4; Gaps 4;

Db 45 PRFVCEVEIPAGPTTSSSITHLRV-EPSTGALL 79
QY 41 PSYV-KFQDNVPGSQTF-SAGL-HLRVDFPSTGALV 73

RESULT 4
ID Q65581 PRELIMINARY; PRT; 300 AA.
AC Q65581;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

```

```

DE UL7 POLYPEPTIDE.
GN UL7.
OS BOVINE HERPESVIRUS TYPE 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.
RN
RP SEQUENCE FROM N.A.
RC STRAIN=SCHOENBOEKEN;
RX MEDLINE: 96135223.
RA SCHWITT J., KEIL G.M.;
RT "Identification and characterization of the bovine herpesvirus 1 UL7
RT gene and gene product which are not essential for virus replication
RT in cell culture.";
RL J. VIROL. 70:1091-1099(1996).
DR EMBL: X91751; G1006630; -.
SQ SEQUENCE 300 AA; 32450 MW; 86377347 CRC32;

Query Match 16.7%; Score 92; DB 14; Length 300;
Best Local Similarity 47.2%; Pred. No. 8.15e-02;
Matches 17; Conservative 9; Mismatches 6; Indels 4; Gaps 4;

Db 46 PRFVCEVEIPAGPTTSSSITHLRV-EPSTGALL 80
QY 41 PSYV-KFQDNVPGSQTF-SAGL-HLRVDFPSTGALV 73

RESULT 5
ID O84354 PRELIMINARY; PRT; 566 AA.
AC O84354;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 63.5 KD PROTEIN.
GN Cr350.
OS CHLAMYDIA TRACHOMATIS.
OC BACTERIA; CHLAMYDIALES; CHLAMYDIACEAE; CHLAMYDIA.
RN
RP SEQUENCE FROM N.A.
RC STRAIN=D/UV-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATSOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT Chlamydia trachomatis.";
RL SCIENCE 0:0-0(1998).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=D/UV-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATSOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AE001308; G3328771; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 566 AA; 63508 MW; C1A1C491 CRC32;

Query Match 16.2%; Score 89; DB 2; Length 566;
Best Local Similarity 26.2%; Pred. No. 2.28e-01;
Matches 17; Conservative 25; Mismatches 18; Indels 5; Gaps 5;

Db 63 IAESYLOQSFSEPTYIR-KSAITGA-GLSGSSEA-LELLSEATQDLYEQL-L-ILNA 117
QY 8 FGCGYVQTPFLSENSVRYKISIAAGSCPLSTAGPSYKVFQDNVPGSQTSFAGLHRLVDFP 67
Db 118 ATSQL 122
QY 68 STGAL 72

RESULT 6
ID Q81862 PRELIMINARY; PRT; 1693 AA.
AC Q81862;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

```

```

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DI ORF 1.
OS HEPATITIS E VIRUS (HEV).
OC VIRUSES: SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALCIVIRIDAE;
OC CALCIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92024067.
RA TAM A.W., SMITH M.M., GUERRA M.E., HUANG C.C., BRADLEY D.W., FRY K.E.,
RA REYES G.R.;
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT full-length viral genome.";
RL VIROLOGY 185:120-131(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92261377.
RA UCHIDA T., SUZUKI K., HAYASHI N., IIDA F., HARA T., OO S.S.,
RA WANG C.K., SHIRATA T., ICHIKAWA M., RIKIHISA T., MIZUNO K., WIN K.M.;
RT "Hepatitis E virus: cDNA cloning and expression.";
RL MICROBIOL. IMMUNOL. 36:67-79(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92335008.
RA AYE T.T., UCHIDA T., MA X.Z., IIDA F., SHIRATA T., ZHUANG H.,
RA WIN K.M.;
RT "Complete nucleotide sequence of a hepatitis E virus isolated from
RT the Xinjiang epidemic (1986-1988) of China.";
RL NUCLEIC ACIDS RES. 20:3512-3512(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92115700.
RA TSAREV S.A., EMERSON S.U., REYES G.R., TSAREVA T.S., LEGTERS L.J.,
RA MALIK I.A., IQBAL M., PURCELL R.H.;
RT "Characterization of a prototype strain of hepatitis E virus.";
RL PROC. NATL. ACAD. SCI. U.S.A. 89:559-563(1992).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92271462.
RA FRY K.E., TAM A.W., SMITH M.M., KIM J.P., LUK K.C., YOUNG L.M.,
RA PIATAK M., FELDMAN R.A., YUN K.Y., PURDY M.A., ET AL.;
RT "Hepatitis E virus (HEV): strain variation in the nonstructural gene
RT region encoding consensus motifs for an RNA-dependent RNA polymerase
RT and an ATP/GTP binding site.";
RL VIRUS GENES 6:173-185(1992).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93348763.
RA BI S.L., PURDY M.A., MCCAUSTLAND K.A., MARGOLIS H.S., BRADLEY D.W.;
RT "The sequence of hepatitis E virus isolated directly from a single
RT source during an outbreak in China.";
RL VIRUS RES. 28:233-247(1993).
DR ENBL: L08816: G33009;
SQ SEQUENCE 1693 AA: 185052 MW: 30A917E1 CRC32;

Query Match 15.6%; Score 86; DB 14; Length 1693;
Best Local Similarity 25.4%; Pred.No. 6.27e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps

Db 572 FRISFVDFGAVLENGPERINLSEFDSQSMTAAGPSLTYAASAGLEVVRYVAAGLDHRV 631
| : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 8 FGQYVQTPFLSESNVRYKIAGSCPLSTAGSPSYVVFQDNPVGQSOT-F-SAGLHLR-V 64

Db 632 FAPGVSP 638
| | : : :

QY 65 FDPSTGA 71

RESULT 7
ID Q81344 PRELIMINARY; PRT: 1693 AA.
AC Q81344:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

```



OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; SIPHOVIRIDAE;  
OC LAMBDA PHAGE GROUP.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94374705.  
RA HARTLEY N.M., MURPHY G.O., BRUTON C.J., CHATER K.F.;  
RT "Sequence of the essential early region of phi C31, a temperate phage  
RI of Streptomyces spp. with unusual features in its lytic  
RT development.";  
RL GENE 147:29-40(1994).  
DR ENBL: X76288: G579071; -.  
SQ SEQUENCE 281 AA; 31680 MW; OE734629 CRC32;

Query Match 15.1%; Score 83; DB 9; Length 281;  
Best Local Similarity 24.1%; Pred.No. 1.68e+00;  
Matches 14; Conservative 17; Mismatches 24; Indels 3; Gaps 3;

Db 194 AYNMADFLIPDGNREPMPEFDGAANVLHVHDETWAFKPVETGPDVFQAFHLHRQTDF 250  
::: | : :::: | : :::: | : : | : : | : : | : : | : : |||| |  
QY 11 GYGVTPELSESVRVKIS-IAGSCPLSTAGPSYRVKFQDNPNVGSTFSAGLHLR-VFD 66

RESULT 14  
ID Q54178 PRELIMINARY; PRT: 1577 AA.  
AC Q54178; Q54247;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DI 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DL 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE GLUCOSYLTRANSFERASE.  
GN GTFG.  
OS STREPTOCOCCUS GORDONII.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
OC STREPTOCOCCUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CHALLIS;  
RX MEDLINE: 96157084.  
RA VICKERMAN M.M., SULAVIK M.C., CLEWELL D.B.;  
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase  
RI phase variants.";  
RL DEV. BIOL. STAND. 85:309-314(1995).  
RN [2]  
RP SEQUENCE OF I-96 FROM N.A.  
RC STRAIN=CHALLIS;  
RX MEDLINE: 92276337.  
RA SULAVIK M.C., TARDIF G., CLEWELL D.B.;  
RT "Identification of a gene, rgg, which regulates expression of  
RI glucosyltransferase and influences the Spp phenotype of Streptococcus  
RT gordonii Challis.";  
RL J. BACTERIOL. 174:3577-3586(1992).  
RR ENBL: U12643: G1054877; -.  
DR ENBL: M89776: G153795; -.  
DR PFAM: PF00128; alpha-amylase; 1.  
KW TRANSFERASE.  
SQ SEQUENCE 1577 AA; 177805 MW; 7D58FEC2 CRC32;

Query Match 15.1%; Score 83; DB 2; Length 1577;  
Best Local Similarity 41.2%; Pred.No. 1.68e+00;  
Matches 14; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Db 239 QDDGTYYKNFAVELNGKILFYDAETGALVDLSNEY 272  
||: | : : | : : | : : | : : | : : | : : ||||| |  
QY 47 QDNPGVSQTFSGAGLHLRV--FDPSTGALVDSKSY 78

RESULT 15  
ID O39221 PRELIMINARY; PRT: 1693 AA.  
AC O39221;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DI 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DL 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE NONSTRUCTURAL POLYPEPTIDE.  
OS HEPATITIS E VIRUS (HEV).



\*\*\*\*\*

W O R L D  
I N F O R M A T I O N  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:29:24 1999; Maspar time 5.26 Seconds  
Tabular output not generated. 145.242 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (66-79) from PCTUS9913024.pep (11 of 12)  
Perfect Score: 86  
Sequence: 1 DPSTGALVDSKSYA 14

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: spiremb19

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 24.488; Variance 26.991; scale 0.907

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	59	68.6	1577	2	GLUCOSYLTRANSFERASE	7.85e-02
2	56	65.1	368	2	HEAT SHOCK PROTEIN/SER	4.07e-01
3	56	65.1	472	2	SERINE PROTEASE.	4.07e-01
4	52	60.5	405	4	KIAA0574 PROTEIN (FRAG	3.35e+00
5	52	60.5	552	5	RADIAL SPOKEHEAD.	3.35e+00
6	52	60.5	870	4	KIAA0563 PROTEIN.	3.35e+00
7	51	59.3	540	5	PHOTOLYASE.	5.58e+00
8	51	59.3	1340	14	POLYPROTEIN (FRAGMENT)	5.58e+00
9	51	59.3	3493	14	POLYPROTEIN.	5.58e+00
10	50	58.1	142	14	ENVELOPE GLYCOPROTEIN	9.22e+00
11	50	58.1	156	11	MONOCARBOXYLATE TRANSP	9.22e+00
12	50	58.1	262	2	W-AMINO-TRANSFERASE-LI	9.22e+00
13	50	58.1	371	2	CMP-N-ACETYLNEURAMINAT	9.22e+00
14	50	58.1	492	11	MONOCARBOXYLATE TRANSP	9.22e+00
15	50	58.1	542	13	RETINAL EPITHELIAL MEM	9.22e+00
16	50	58.1	542	13	SIMILARITY TO MYOSIN H	9.22e+00
17	50	58.1	1133	10	PATERNALLY EXPRESSED P	9.22e+00
18	50	58.1	1378	11	YCDH.	1.51e+01
19	49	57.0	319	2	HEMAGGLUTININ (FRAGMEN	1.51e+01
20	49	57.0	327	14	HEMAGGLUTININ (FRAGMEN	1.51e+01

21	49	57.0	329	14	HEMAGGLUTININ (FRAGMEN	1.51e+01
22	49	57.0	329	14	HEMAGGLUTININ (FRAGMEN	1.51e+01
23	49	57.0	427	5	PARASITOPHOROUS VACUOL	1.51e+01
24	49	57.0	734	11	INSULIN RECEPTOR SUBST	1.51e+01
25	49	57.0	966	1	HYPOTHETICAL 104.9 KD	1.51e+01
26	49	57.0	984	5	SERINE RICH PROTEIN (S	1.51e+01
27	49	57.0	1139	2	ANCHOR PROTEIN, LCM.	1.51e+01
28	48	55.8	195	2	HYPOTHETICAL 20.8 KD P	2.46e+01
29	48	55.8	328	14	HAEMAGGLUTININ (HAI DO	2.46e+01
30	48	55.8	329	14	HEMAGGLUTININ (FRAGMEN	2.46e+01
31	48	55.8	329	14	HEMAGGLUTININ (FRAGMEN	2.46e+01
32	48	55.8	329	14	HEMAGGLUTININ (FRAGMEN	2.46e+01
33	48	55.8	329	14	HEMAGGLUTININ (FRAGMEN	2.46e+01
34	48	55.8	329	14	HEMAGGLUTININ (FRAGMEN	2.46e+01
35	48	55.8	329	14	HEMAGGLUTININ (FRAGMEN	2.46e+01
36	48	55.8	427	10	S-LOCUS-SPECIFIC GLYCO	2.46e+01
37	48	55.8	477	11	SCK, PARTIAL CDS (FRAG	2.46e+01
38	48	55.8	480	2	PUTATIVE TRANSCRIPTION	2.46e+01
39	48	55.8	540	4	SCK, PARTIAL CDS (FRAG	2.46e+01
40	48	55.8	550	14	HAEMAGGLUTININ PRECURS	2.46e+01
41	48	55.8	580	3	DIHYDROXYACETONE KINAS	2.46e+01
42	48	55.8	712	2	OMP1D PRECURSOR.	2.46e+01
43	48	55.8	1571	11	ZINC FINGER PROTEIN.	2.46e+01
44	47	54.7	329	14	HEMAGGLUTININ (FRAGMEN	3.97e+01
45	47	54.7	329	14	HEMAGGLUTININ (FRAGMEN	3.97e+01

ALIGNMENTS

RESULT	1					
ID	Q54178	PRELIMINARY;	PRT:	1577	AA.	
AC	Q54178; Q54247;					
DT	01-NOV-1996 (TREMBREL.	01, CREATED)				
DT	01-NOV-1996 (TREMBREL.	01, LAST SEQUENCE UPDATE)				
DT	01-NOV-1998 (TREMBREL.	08, LAST ANNOTATION UPDATE)				
DE	GLUCOSYLTRANSFERASE.					
GN	GTFG.					
OS	STREPTOCOCCUS GORDONII.					
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;					
OC	STREPTOCOCCUS.					
RN	[1]					
RC	SEQUENCE FROM N.A.					
RC	STRAIN-CHALLIS;					
RX	MEDLINE: 96157084.					
RA	VICKERMAN M.M., SULAVIK M.C., CLEWELL D.B.;					
RT	"Molecular analysis of Streptococcus gordonii glucosyltransferase					
RT	phase variants."					
RL	DEV. BIOL. STAND. 85:309-314(1995).					
RN	[2]					
RP	SEQUENCE OF 1-96 FROM N.A.					
RC	STRAIN-CHALLIS;					
RX	MEDLINE: 92276337.					
RA	SULAVIK M.C., TARDIF G., CLEWELL D.B.;					
RT	"Identification of a gene, rgg, which regulates expression of					
RT	glucosyltransferase and influences the Spp phenotype of Streptococcus					
RL	J. BACTERIOL. 174:3577-3586(1992).					
DR	EMBL: U12643; G1054877; -					
DR	EMBL: M89776; G133795; -					
DR	PFAM: PF00128; alpha-amylase; 1.					
KW	TRANSFERASE.					
SQ	SEQUENCE 1577 AA; 177805 MW; 7D58FEC2 CRC32;					
	Query Match	58.6%;	Score	59;	DB 2: Length 1577;	
	Best Local Similarity	69.2%;	Pred. No.	7.85e-02;		
	Matches	9;	Conservative	2;	Mismatches	0;
					Indels	0;
DB	260	DAETGALVDSNEY	272			
QY	66	DPSTGALVDSKSY	78			
		1:				
RESULT	2					

ID Q46094 PRELIMINARY; PRT; 368 AA.  
 AC Q46094;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE HEAT SHOCK PROTEIN/SERINE PROTEASE (FRAGMENT).  
 GN HTRA.  
 OS CAMPYLOBACTER JEJUNI.  
 OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; CAMPYLOBACTER GROUP;  
 OC CAMPYLOBACTER.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-UA580;  
 RX MEDLINE; 90384493.  
 RA TAYLOR D.E., HIRATSUKA K.;  
 RT "Use of non-radioactive DNA probes for detection of Campylobacter  
 RT jejuni and Campylobacter coli in stool specimens.";  
 RL MOL. CELL. PROBES 4:261-271(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-UA580;  
 RA HIRATSUKA K.;  
 RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; U27271; G881375;  
 DR PFAM; PF00089; trypsin; 1.  
 DR PFAM; PF00595; PDZ; 2.  
 KW HEAT SHOCK; PROTEASE.  
 FT NON\_TER 1  
 SQ SEQUENCE 368 AA; 39491 MW; A5062589 CRC32;

Query Match 65.1%; Score 56; DB 2; Length 368;  
 Best Local Similarity 72.7%; Pred. No. 4.07e-01;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 121 SGGALVDSRGY 131  
 I | | | | | | | | | |

QY 68 STGALVDSKSY 78  
 I | | | | | | | | | |

RESULT 3  
 ID Q46120 PRELIMINARY; PRT; 472 AA.  
 AC Q46120;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE SERINE PROTEASE.  
 GN HTRA.  
 OS CAMPYLOBACTER JEJUNI.  
 OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; CAMPYLOBACTER GROUP;  
 OC CAMPYLOBACTER.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-81116;  
 RA HENDERSON J., WOOD A.C., EMERY M.J., WREN B.W., KETLEY J.;  
 RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; X82628; E315368;  
 DR PFAM; PF00089; trypsin; 1.  
 DR PFAM; PF00595; PDZ; 2.  
 KW PROTEASE.  
 SQ SEQUENCE 472 AA; 50940 MW; 6BCA101 CRC32;

Query Match 65.1%; Score 56; DB 2; Length 472;  
 Best Local Similarity 72.7%; Pred. No. 4.07e-01;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 225 SGGALVDSRGY 235  
 I | | | | | | | | | |

QY 68 STGALVDSKSY 78  
 I | | | | | | | | | |

RESULT 4  
 ID Q60320 PRELIMINARY; PRT; 405 AA.  
 AC Q60320;

DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE KIAA0574 PROTEIN (FRAGMENT).  
 GN KIAA0574.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE; 98290545.  
 RA NAGASE T., ISHIKAWA K., MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N.,  
 RA OHARA O.;  
 RT "Prediction of the coding sequences of unidentified human genes. IX.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA RES. 5:31-39(1998).  
 DR EMBL; AB011146; D1026430;  
 FT NON\_TER 1  
 SQ SEQUENCE 405 AA; 42409 MW; 09BA9932 CRC32;

Query Match 60.5%; Score 52; DB 4; Length 405;  
 Best Local Similarity 77.8%; Pred. No. 3.35e+00;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 345 SLVDSKAYA 353  
 : | | | | | | | |

QY 71 ALVDSKSYA 79  
 : | | | | | | | |

RESULT 5  
 ID Q46178 PRELIMINARY; PRT; 552 AA.  
 AC Q46178;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE RADIAL SPOKEHEAD.  
 OS STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).  
 OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; EUECHINOIDEA;  
 OC ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE; STRONGYLOCENTROTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA GINGRAS D., GAGNON C.;  
 RT "Molecular cloning and characterization of a radial spoke head  
 RT protein of sea urchin sperm axonemes: involvement of the protein in  
 RT the regulation of sperm motility.";  
 RL MOL. BIOL. CELL 9:513-522(1998).  
 DR EMBL; U73123; G2905895;  
 SQ SEQUENCE 552 AA; 62723 MW; 898CFCCC CRC32;

Query Match 60.5%; Score 52; DB 5; Length 552;  
 Best Local Similarity 50.0%; Pred. No. 3.35e+00;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 39 PTEALVNKAY 50  
 I : | | | | | | |

QY 67 PSTGALVDSKSY 78  
 I : | | | | | | |

RESULT 6  
 ID Q60309 PRELIMINARY; PRT; 870 AA.  
 AC Q60309;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE KIAA0563 PROTEIN.  
 GN KIAA0563.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]



```
Oy 66 DPSTGALVDSKSYA 79
||| |||::| :|:
RESULT 10 PRELIMINARY; PRT; 142 AA.
ID Q87600;
AC Q87600;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS SIMIAN IMMUNODEFICIENCY VIRUS (SIV-AGM).
OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; LENTIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P081.
RA BIBOLLET-RUCHE F., BRENGUES C., GALAT-LUONG A., GALAT G., POURRUT X.,
RA VIDAL N., YEAS F., DURAND J.P., CUNY G.;
RL J. VIROL. 71:307-313(1997).
DR EMBL; U37209; G1353456; -.
DR PFAM; PF00517; GP41; 1.
KW ENVELOPE PROTEIN.
FT NON_TER 1 1
FT NON_TER 142 142
SQ SEQUENCE 142 AA; 16852 MW; C58A9A91 CRC32;

Query Match 58.1%; Score 50; DB 14; Length 142;
Best Local Similarity 46.7%; Pred. No. 9.22e+00;
Matches 7; Conservative 5; Mismatches 2; Indels 1; Gaps 1;

Db 86 EPDSGGLSRDSRYT 100
:|: |||: |||:
Oy 66 DPSTGALV-DSKSYA 79

RESULT 11 PRELIMINARY; PRT; 156 AA.
ID Q35308;
AC Q35308;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE MONOCARBOXYLATE TRANSPORTER MCT3 (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H;
RA PHILP N.J., YOON H.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF019111; G2407664; -.
FT NON_TER 1 1
SQ SEQUENCE 156 AA; 16253 MW; DB394B9E CRC32;

Query Match 58.1%; Score 50; DB 11; Length 156;
Best Local Similarity 61.5%; Pred. No. 9.22e+00;
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 37 PSAGRLVDALKNY 49
||: |||: |||:
Oy 67 PSTGALVDS-KSY 78

RESULT 12 PRELIMINARY; PRT; 262 AA.
ID Q44188;
AC Q44188;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE W-AMINO-TRANSFERASE-LIKE PROTEIN.
GN OATA.
OS AGROBACTERIUM RADIOBACTER.
```

```
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; AGROBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fl305 LAC9;
RX MEDLINE; 96425889.
RA TIBURTIUS A., DE LUCA N., HUSSAIN H., JOHNSTON A.W.B.;
RT "Expression of the exy gene, required for exopolysaccharide
RT synthesis in Agrobacterium, is activated by the regulatory ros
RT gene.";
RL MICROBIOLOGY 142:2621-2629(1996).
DR EMBL; X95394; E220383; -.
DR PFAM; PF00202; aminotran_3; 1.
KW TRANSFERASE.
SQ SEQUENCE 262 AA; 28124 MW; DF7A4568 CRC32;

Query Match 58.1%; Score 50; DB 2; Length 262;
Best Local Similarity 63.6%; Pred. No. 9.22e+00;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 110 STGVLVPPKGY 120
||| |||: |||:
Oy 68 STGALVDSKSY 78

RESULT 13 PRELIMINARY; PRT; 371 AA.
ID P72097;
AC P72097; P72100; P72101;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CMP-N-ACETYLNEURAMINATE-BETA-GALACTOSAMIDE-ALPHA-2,3-
DE SIALYLTRANSFERASE (EC 2.4.99.-)
DE (BETA-GALACTOSIDE ALPHA-2,3-SIALYLTRANSFERASE) (ALPHA 2,3-ST)
DE (LIPOOLIGOSACCHARIDE SIALYLTRANSFERASE) (LST).
GN LST.
OS NEISSERIA MENINGITIDIS.
OC BACTERIA; PROTEOBACTERIA; BETA SUBDIVISION; NEISSERIA;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / NRCC 4728, 406Y / NRCC 4030, AND M982B / NRCC 4725;
RA GILBERT M., WATSON D.C., CUNNINGHAM A.-M., JENNINGS M.P., YOUNG N.M.,
RA MAKARCHUK W.W.;
RL J. BIOL. CHEM. 271:28271-28276(1996).
CC -1- FUNCTION: TRANSFERS SIALIC ACID FROM THE SUBSTRATE CMP-SIALIC
CC ACID DONOR TO THE TERMINAL
CC BETA-D-GALACTOSYL-1,4-ACETYL-BETA-D-GLUCOSAMINE ON THE
CC LACTO-N-NEOTETRAOSE BRANCH OF THE LIPOOLIGOSACCHARIDE.
CC -1- CATALYTIC ACTIVITY: CMP-N-ACETYLNEURAMINATE +
CC BETA-D-GALACTOSYL-1,4-ACETYL-BETA-D-GLUCOSAMINE = CMP +
CC ALPHA-N-ACETYLNEURAMINYL-2,3-BETA-D-GALACTOSYL-1,4-N-ACETYL-BETA-
CC D-GLUCOSAMINE
CC -1- PATHWAY: GLYCOSYLATION.
DR EMBL; U60660; G1546004; -.
DR EMBL; U60661; G1546007; -.
DR EMBL; U60662; G1546009; -.
DR EMBL; U60663; G1546011; -.
KW TRANSFERASE; GLYCOSYLTRANSFERASE.
FT VARIANT 2 2 G -> S (IN STRAIN M982B / NRCC 4725).
FT VARIANT 29 29 Q -> H
FT (IN STRAINS 406Y / NRCC 4030 AND M982B /
FT NRCC 4725).
FT E -> D
FT (IN STRAINS 406Y / NRCC 4030 AND M982B /
FT NRCC 4725).
FT N -> K
FT (IN STRAINS 406Y / NRCC 4030 AND M982B /
FT NRCC 4725).
FT L -> *: (IN STRAIN M982B / NRCC 4725;
FT LOSS OF ACTIVITY).
FT T -> A (IN STRAIN 406Y / NRCC 4030).
FT K -> N (IN STRAIN 406Y / NRCC 4030).
FT SEQUENCE 371 AA; 42611 MW; 73DB83A9 CRC32;
```

Query Match 58.1%; Score 50; DB 2; Length 371;  
Best Local Similarity 53.8%; Pred. No. 9.22e+00;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 164 DGGTGNLIQSSSY 176  
QY 66 DPSTGALVDSKSY 78  
I:||||:|I|

RESULT 14  
ID 070461 PRELIMINARY; PRT; 492 AA.  
AC 070461;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE MONOCARBOXYLATE TRANSPORTER MCT3.  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-RETINA;  
RA PHILP N.J., YOON H.;  
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR ENBL; AF059258; G3170609; -  
SQ SEQUENCE 492 AA; 51590 MW; 20F051C1 CRC32;

Query Match 58.1%; Score 50; DB 11; Length 492;  
Best Local Similarity 61.5%; Pred. No. 9.22e+00;  
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 373 PSAGRLVDALKNY 385  
QY 67 PSTGALVDS-KSY 78  
I:||||:|I|

RESULT 15  
ID 013151 PRELIMINARY; PRT; 542 AA.  
AC 013151;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE MONOCARBOXYLATE TRANSPORTER 3.  
GN MCT3.  
OS GALLUS GALLUS (CHICKEN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LEGHORN;  
RA PHILP N.J., YOON H.;  
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR ENBL; AF000240; G2198807; -  
SQ SEQUENCE 542 AA; 58085 MW; 836DA7B1 CRC32;

Query Match 58.1%; Score 50; DB 13; Length 542;  
Best Local Similarity 61.5%; Pred. No. 9.22e+00;  
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 410 PSAGRLVDALKNY 422  
QY 67 PSTGALVDS-KSY 78  
I:||||:|I|

Search completed: Wed Sep 1 16:29:45 1999  
Job time : 21 secs.

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:33:07 1999; MasPar time 4.43 Seconds  
62.340 Million cell updates/sec

Tabular output not generated.

Title: >PCT-US99-13024-2

Description: (81-93) from PCTUS9913024.pep (12 of 12)

Perfect Score: 68

Sequence: 1 STSNDTSAAFVS 13

Scoring table: PAM 150

Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 15.573; Variance 46.313; scale 0.336

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	54	79.4	216	33	Human calcium channel	1.78e+01
2	54	79.4	216	14	Human calcium channel	1.78e+01
3	54	79.4	219	14	Human calcium channel	1.78e+01
4	54	79.4	219	33	Human calcium channel	1.78e+01
5	54	79.4	240	8	Myasthenic antigenic	1.78e+01
6	54	79.4	478	8	Human neuronal VDCC b	1.78e+01
7	54	79.4	478	33	Human calcium channel	1.78e+01
8	54	79.4	478	14	Human calcium channel	1.78e+01
9	54	79.4	478	6	Sequence of splice va	1.78e+01
10	54	79.4	479	14	Human calcium channel	1.78e+01
11	54	79.4	523	33	Human calcium channel	1.78e+01
12	54	79.4	530	33	Human calcium channel	1.78e+01
13	54	79.4	571	8	Human neuronal VDCC b	1.78e+01
14	54	79.4	598	33	Human calcium channel	1.78e+01
15	54	79.4	598	14	Human calcium channel	1.78e+01
16	54	79.4	598	6	Sequence of splice va	1.78e+01

17	54	79.4	660	14	R72613	Human neuronal calciu	1.78e+01
18	44	64.7	636	36	W79589	Human kv potassium ch	2.29e+02
19	44	64.7	636	36	W79591	Human kv potassium ch	2.29e+02
20	44	64.7	655	36	W79590	Human kv potassium ch	2.29e+02
21	44	64.7	1130	15	R71729	Merosin major subunit	2.29e+02
22	44	64.7	1130	3	R13436	Merosin M polypeptide	2.29e+02
23	44	64.7	3110	15	R71730	Merosin major subunit	2.29e+02
24	43	63.2	757	20	W03179	Bovine poly-immunoglo	2.93e+02
25	42	61.8	127	28	W31716	Gamma-1 heavy chain a	3.74e+02
26	42	61.8	653	21	W14264	Z. japonica phosphoen	3.74e+02
27	42	61.8	953	19	R97284	Human 26S proteasome	3.74e+02
28	42	61.8	2783	5	R23963	AFP-1 (Ala 2460 Val)	3.74e+02
29	42	61.8	2783	5	R23962	AFP-1.	3.74e+02
30	41	60.3	98	13	R72068	Dp10 VH region.	4.76e+02
31	41	60.3	120	25	W27550	Human Ab heavy chain	4.76e+02
32	41	60.3	120	10	R54795	SpA-reactive IgM heav	4.76e+02
33	41	60.3	123	24	W19881	CEA-specific antibody	4.76e+02
34	41	60.3	123	24	W19889	CEA-specific antibody	4.76e+02
35	41	60.3	123	24	W19888	CEA-specific antibody	4.76e+02
36	41	60.3	123	24	W19887	CEA-specific antibody	4.76e+02
37	41	60.3	128	28	W31715	Gamma-1 heavy chain a	4.76e+02
38	41	60.3	268	27	W27135	Achromobacter lyticus	4.76e+02
39	41	60.3	268	27	W31403	Achromobacter lyticus	4.76e+02
40	41	60.3	268	26	W27135	Achromobacter lyticus	4.76e+02
41	41	60.3	397	19	W04270	B.t. alkaline protease	4.76e+02
42	41	60.3	481	5	R24442	Sequence of antibody	4.76e+02
43	41	60.3	512	39	W87797	Protease encoded by c	4.76e+02
44	41	60.3	512	14	R80505	S. lividans protease	4.76e+02
45	41	60.3	1170	12	R63231	Crystal protein CryET	4.76e+02

## ALIGNMENTS

RESULT 1

ID W63157 standard; Protein; 216 AA.

AC W63157;

DT 12-OCT-1998 (first entry)

DE Human calcium channel beta-4 subunit.

KW Beta-4 subunit; human; calcium channel; assay; detection;

OS Characterisation; Lambert Eaton Syndrome; LES; diagnosis.

PN Homo sapiens.

PD US792846-A.

PF 11-AUG-1998.

PR 31-MAY-1995; 455543.

PR 04-APR-1994; US-223305.

PR 04-APR-1988; US-176899.

PR 04-APR-1989; US-603751.

PR 04-APR-1989; WO-001408.

PR 20-FEB-1990; US-482384.

PR 30-NOV-1990; US-620250.

PR 15-AUG-1991; US-745206.

PR 31-MAY-1995; US-455543.

PA (SIBI-) SIBIA NEUROSCIENCES INC.

PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,

PI Williams ME;

DR WPI; 98-456192/39.

DR N-PSDB: V42706.

PT DNA encoding human calcium channel alpha 1B subunit protein -

its modulators, and diagnosis of Lambert Eaton Syndrome

FS Disclosure; Columns 225-228; 16pp; English.

CC The present sequence represents the beta-4 subunit of a human calcium

channel. Calcium channels are membrane-spanning, multi-subunit proteins

that allow controlled entry of calcium ions into cells. This leads

to depolarisation events required for muscle contraction. The recombinant

subunit, when expressed with nucleic acids encoding the complete calcium

channel, can be used in assays for the detection and characterisation of

compounds that modulate the channel. The DNA encoding the subunits can

be alternatively spliced when transcribed, giving more than one form of

the protein from the same transcript, each having slightly different

properties. In addition, the reactivity of the alpha 1 subunit with IgG

molecules from the serum of an individual with Lambert Eaton Syndrome

(LES) can be used as a diagnostic for the disease.

```

SQ Sequence 216 AA;
Query Match 79.4%; Score 54; DB 33; Length 216;
Best Local Similarity 75.0%; Pred. No. 1.78e+01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnf 55
   |||:||||: ||
QY 81 STSNDTTSAAFY 92

RESULT 2
ID R72605 standard; Protein; 216 AA.
AC R72605;
DE 01-DEC-1995 (first entry)
DT Human calcium channel subunit beta 1-4.
KW Calcium channel subunit; antagonist; agonist; diagnosis;
KW Lambert Eaton Syndrome.
OS Homo sapiens.
PN W09504822-A.
PD 16-FEB-1995.
PF 11-AUG-1994; U09230.
PR 11-AUG-1993; US-105536.
PR 05-NOV-1993; US-149097.
PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.
PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;
DR N-PSDB; Q87832.
DR N-PSDB; Q87832.
PT DNA encoding human calcium channel sub-unit(s) - used for
PT developing prods. for studying calcium channels, e.g. for
PT obtaining agonists and antagonists
PS Disclosure: Page 256-257; 285pp; English.
CC To isolate DNA encoding the Beta 1 subunit, a human hippocampus
CC cDNA library was screened by hybridisation to a DNA fragment
CC encoding a rabbit skeletal muscle calcium channel beta subunit.
CC A hybridising clone was selected and was in turn used to isolate
CC overlapping clones until the overlapping clones encompassing DNA
CC encoding the entire human calcium channel beta 2 subunit were
CC isolated and sequenced. Five alternatively spliced forms of the
CC beta 1 subunit have been identified. These forms are designated
CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in the
CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in
CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293
CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3
CC subunits have been constructed. The subunits beta 1-1, beta 1-2,
CC beta 1-4 and beta 1-5 have been identified by nucleic acid
CC amplification analysis as alternatively spliced forms of the
CC beta subunit. Sequences of the beta 1 splice variants are set
CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and
CC R72604-R72706.
SQ Sequence 216 AA;

Query Match 79.4%; Score 54; DB 14; Length 216;
Best Local Similarity 75.0%; Pred. No. 1.78e+01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnf 55
   |||:||||: ||
QY 81 STSNDTTSAAFY 92

RESULT 3
ID R72606 standard; Protein; 219 AA.
AC R72606;
DE 01-DEC-1995 (first entry)
DT Human calcium channel subunit beta 1-5.
KW Calcium channel subunit; antagonist; agonist; diagnosis;
KW Lambert Eaton Syndrome.
OS Homo sapiens.
PN W09504822-A.
PD 16-FEB-1995.
PF 11-AUG-1994; U09230.
PR 11-AUG-1993; US-105536.

```

```

PR 05-NOV-1993; US-149097.
PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.
PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;
DR WPI: 95-090900/12.
DR N-PSDB; Q87833.
PT DNA encoding human calcium channel sub-unit(s) - used for
PT developing prods. for studying calcium channels, e.g. for
PT obtaining agonists and antagonists
PS Disclosure: Page 258-259; 285pp; English.
CC To isolate DNA encoding the Beta 1 subunit, a human hippocampus
CC cDNA library was screened by hybridisation to a DNA fragment
CC encoding a rabbit skeletal muscle calcium channel beta subunit.
CC A hybridising clone was selected and was in turn used to isolate
CC overlapping clones until the overlapping clones encompassing DNA
CC encoding the entire human calcium channel beta 2 subunit were
CC isolated and sequenced. Five alternatively spliced forms of the
CC beta 1 subunit have been identified. These forms are designated
CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in the
CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in
CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293
CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3
CC subunits have been constructed. The subunits beta 1-1, beta 1-2,
CC beta 1-4 and beta 1-5 have been identified by nucleic acid
CC amplification analysis as alternatively spliced forms of the
CC beta subunit. Sequences of the beta 1 splice variants are set
CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and
CC R72604-R72706.
SQ Sequence 219 AA;

Query Match 79.4%; Score 54; DB 14; Length 219;
Best Local Similarity 75.0%; Pred. No. 1.78e+01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnf 55
   |||:||||: ||
QY 81 STSNDTTSAAFY 92

RESULT 4
ID W63158 standard; Protein; 219 AA.
AC W63158;
DE 12-OCT-1998 (first entry)
DT Human calcium channel beta-5 subunit.
KW Beta-5 subunit; human; calcium channel; assay; detection; ds.
KW Characterisation; Lambert Eaton Syndrome; LES; diagnosis; ds.
OS Homo sapiens.
PN US5792846-A.
PD 11-AUG-1998.
PF 31-MAY-1995; 455543.
PR 04-APR-1994; US-223305.
PR 04-APR-1988; US-176899.
PR 04-APR-1989; US-603751.
PR 04-APR-1989; WO-001408.
PR 20-FEB-1990; US-482384.
PR 30-NOV-1990; US-620250.
PR 15-AUG-1991; US-745206.
PR 31-MAY-1995; US-455543.
PA (SIBI-) SIBIA NEUROSCIENCES INC.
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
PI Williams ME;
DR WPI: 98-456192/39.
DR N-PSDB; V42707.
PT DNA encoding human calcium channel alpha 1B subunit protein -
PT useful for recombinant production of the channel for screening of
PT its modulators, and diagnosis of Lambert Eaton Syndrome
PS Disclosure: Columns 227-230; 166pp; English.
CC The present sequence represents the beta-5 subunit of a human calcium
CC channel. Calcium channels are membrane-spanning, multi-subunit proteins
CC that allow controlled entry of calcium ions into cells. This leads
CC to depolarisation events required for muscle contraction. The recombinant
CC subunit, when expressed with nucleic acids encoding the complete calcium
CC channel, can be used in assays for the detection and characterisation of
CC compounds that modulate the channel. The DNA encoding the subunits can

```

CC be alternatively spliced when transcribed, giving more than one form of  
 CC the protein from the same transcript, each having slightly different  
 CC properties. In addition, the reactivity of the alpha 1 subunit with IgG  
 CC molecules from the serum of an individual with Lambert Eaton Syndrome  
 CC (LES) can be used as a diagnostic for the disease.  
 SQ Sequence 219 AA;

Query Match 79.4%; Score 54; DB 33; Length 219;  
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfv 55  
 |||:||||: ||  
 QY 81 STSNDTTSAAFY 92

## RESULT 5

ID R39697 standard; Protein; 240 AA.  
 AC R39697;  
 DT 19-JAN-1994 (first entry)  
 DE Myasthenic antigenic polypeptide.  
 KW MAP; antibody detection; LEMS; Lambert-Eaton myasthenic syndrome;  
 KW paraneoplastic sensory neuropathy; p mys B; mysB; ss.  
 OS Homo sapiens.  
 PN WO9314098-A.  
 PD 22-JUL-1993.  
 PF 11-JAN-1993; U00227.  
 PR 10-JAN-1992; US-820312.  
 PA (SLOK ) SLOAN KETTERING INST CANCER.  
 PI Furneaux HM, Posner JB;  
 DR WPI: 93-243126/30.  
 DR N-PSDB: Q46673.  
 PT New purified myasthenic antigenic polypeptide and its corresp.  
 PT antibody - useful for diagnosing and treating proliferation of  
 PT neoplastic cells in patient with Lambert-Eaton myasthenic syndrome  
 PS Claim 5; Page 26-27; 48pp; English.  
 CC The sequence is that of myasthenic antigenic polypeptide (MAP)  
 CC encoded by the cDNA clone p mysB. MAP can be used to detect  
 CC antibodies associated with paraneoplastic sensory neuropathy  
 CC such as Lambert-Eaton myasthenic syndrome (LEMS). These antibodies  
 CC are used to determine if a patient with neurological symptoms has a  
 CC tumour expressing MAP. To inhibit proliferation of neoplastic cells  
 CC in patients with LEMS and for imaging neoplastic cells in LEMS  
 CC patients.  
 SQ Sequence 240 AA;

Query Match 79.4%; Score 54; DB 8; Length 240;  
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 17 stsdtttsnfv 28  
 |||:||||: ||  
 QY 81 STSNDTTSAAFY 92

## RESULT 6

ID R39563 standard; Protein; 478 AA.  
 AC R39563;  
 DT 09-FEB-1994 (first entry)  
 DE Human neuronal VDCC beta-subunit encoded by clone HBBI.  
 KW Voltage-dependent calcium channel; VDCC; beta-subunit;  
 KW calcium flux; ss.  
 OS Homo sapiens.  
 PN DE422126-A.  
 PD 19-AUG-1993.  
 PF 06-JUL-1992; 222126.  
 PR 17-FEB-1992; DE-204716.  
 PR 06-JUL-1992; DE-222126.  
 PA (FARB ) BAYER AG.  
 PI Spreyer P, Unterbeck A;  
 DR WPI: 93-265734/34.  
 DR N-PSDB: Q46076.  
 PT Human neuronal beta-unit cDNA of voltage dependent calcium

PT channels - useful in calcium-flux studies and screening systems  
 PT for agonists and antagonists of calcium channels  
 PS Claim 2; Page 5-7; 13pp; German.  
 CC A first oligonucleotide probe (Q46075) complementary to nucleotides  
 CC 361-400 of the VDCC beta-subunit from rabbit skeletal muscle was  
 CC used to screen a human hippocampus cDNA library. A 1.9kb cDNA  
 CC fragment was isolated for further screening to isolate human VDCC  
 CC beta-subunit coding sequences. Clone HBBI was sequenced and found  
 CC to have 92% homology to the rabbit beta-subunit. The amino acid  
 CC sequence R39563 was deduced from the open reading frame.  
 CC See also Q46077-Q46078.  
 SQ Sequence 478 AA;

Query Match 79.4%; Score 54; DB 8; Length 478;  
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfv 55  
 |||:||||: ||  
 QY 81 STSNDTTSAAFY 92

## RESULT 7

ID W63143 standard; Protein; 478 AA.  
 AC W63143;  
 DT 12-OCT-1998 (first entry)  
 DE Human calcium channel beta subunit splice variant beta2.  
 KW Beta subunit; human; calcium channel; assay; detection;  
 KW characterisation; Lambert Eaton Syndrome; LES; diagnosis.  
 OS Homo sapiens.  
 PN US5792846-A.  
 PD 11-AUG-1998.  
 PF 31-MAY-1995; 455543.  
 PR 04-APR-1994; US-223305.  
 PR 04-APR-1988; US-176899.  
 PR 04-APR-1989; US-603751.  
 PR 04-APR-1989; WO-U01408.  
 PR 20-FEB-1990; US-482384.  
 PR 30-NOV-1990; US-620250.  
 PR 15-AUG-1991; US-745206.  
 PR 31-MAY-1995; US-455543.  
 PA (SIBI-) SIBIA NEUROSCIENCES INC.  
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,  
 PI Williams ME;  
 DR WPI: 98-456192/39.  
 DR N-PSDB: V42687.  
 PT DNA encoding human calcium channel alpha 1B subunit protein -  
 PT useful for recombinant production of the channel for screening of  
 PT its modulators, and diagnosis of Lambert Eaton Syndrome  
 PS Claim 3; Columns 217-220; 166pp; English.  
 CC The present sequence is encoded by a splice variant of the beta subunit  
 CC of a human calcium channel. Calcium channels are membrane-spanning,  
 CC multi-subunit proteins that allow controlled entry of calcium ions into  
 CC cells. This leads to depolarisation events required for muscle  
 CC contraction. The recombinant subunit, when expressed with nucleic acids  
 CC encoding the complete calcium channel, can be used in assays for the  
 CC detection and characterisation of compounds that modulate the channel.  
 CC The DNA encoding the subunits can be alternatively spliced when  
 CC transcribed, giving more than one form of the protein from the same  
 CC transcript, each having slightly different properties. In addition, the  
 CC reactivity of the alpha 1 subunit with IgG molecules from the serum of  
 CC an individual with Lambert Eaton Syndrome (LES) can be used as a  
 CC diagnostic for the disease.  
 SQ Sequence 478 AA;

Query Match 79.4%; Score 54; DB 33; Length 478;  
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfv 55  
 |||:||||: ||  
 QY 81 STSNDTTSAAFY 92



RESULT 8  
 ID R72611 standard; Protein; 478 AA.  
 AC R72611:  
 DT 01-DEC-1995 (first entry)  
 DE Human calcium channel subunit beta 1.  
 KW Calcium channel subunit; antagonist; agonist; diagnosis;  
 KW Lambert Eaton Syndrome.  
 OS Homo sapiens.  
 PN W09504822-A.  
 PD 16-FEB-1995.  
 PF 11-AUG-1994; U09230.  
 PR 11-AUG-1993; US-105536.  
 PR 05-NOV-1993; US-149097.  
 PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.  
 PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;  
 DR WPI: 95-090900/12.  
 DR N-PSDB: Q87838.  
 PT DNA encoding human calcium channel sub-unit(s) - used for  
 PT developing prods. for studying calcium channels, e.g. for  
 PT obtaining agonists and antagonists  
 PS Disclosure: Page 160-162; 285pp; English.  
 CC To isolate DNA encoding the Beta 1 subunit, a human hippocampus  
 CC cDNA library was screened by hybridisation to a DNA fragment  
 CC encoding a rabbit skeletal muscle calcium channel beta subunit.  
 CC A hybridising clone was selected and was in turn used to isolate  
 CC overlapping clones until the overlapping clones encompassing DNA  
 CC encoding the entire human calcium channel beta 2 subunit were  
 CC isolated and sequenced. Five alternatively spliced forms of the  
 CC beta 1 subunit have been identified. These forms are designated  
 CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in the  
 CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in  
 CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293  
 CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3  
 CC subunits have been constructed. The subunits beta 1-1, beta 1-2,  
 CC beta 1-4 and beta 1-5 have been identified by nucleic acid  
 CC amplification analysis as alternatively spliced forms of the  
 CC beta subunit. Sequences of the beta 1 splice variants are set  
 CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and  
 CC R72604-R72706.  
 SQ Sequence 478 AA;  
 Query Match 79.4%; Score 54; DB 14; Length 478;  
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 44 stsdtsnsfv 55  
 QY 81 STSNDTTSAAVF 92  
 III:IIII:II  
 RESULT 9  
 ID R33551 standard; Protein; 478 AA.  
 AC R33551:  
 DT 30-JUN-1993 (first entry)  
 DE Sequence of splice variant beta 1-2 of beta human calcium  
 DE channel subunit.  
 KW Human calcium channel subunit; diagnosis; agonist; antagonist;  
 KW Lambert Eaton syndrome.  
 OS Homo sapiens.  
 PN W09304083-A.  
 PD 04-MAR-1993.  
 PF 14-AUG-1992; U06903.  
 PR 15-AUG-1991; US-745206.  
 PR 10-APR-1992; US-868354.  
 PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.  
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,  
 PI Williams ME;  
 DR WPI: 93-093936/11.  
 DR N-PSDB: Q37819.  
 PT DNA encoding specific human calcium channel sub-units - used for  
 PT identifying calcium channel agonists and antagonists and  
 PT diagnosing Lambert Eaton syndrome

PS Disclosure: Page 129-131; 150pp; English.  
 CC Five alternatively spliced forms of the human calcium channel  
 CC beta 1 subunit have been identified and DNA encoding a number  
 CC of forms have been isolated. These forms are designated beta 1-1,  
 CC expressed in skeletal muscle, beta 1-2, expressed in the CNS, beta  
 CC 1-3, also expressed in the CNS, beta 1-4, expressed in aorta tissue  
 CC and HEK 293 cells, and beta 1-5, expressed in HEK 293 cells.  
 SQ Sequence 478 AA;  
 Query Match 79.4%; Score 54; DB 6; Length 478;  
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 44 stsdtsnsfv 55  
 QY 81 STSNDTTSAAVF 92  
 III:IIII:II  
 RESULT 10  
 ID R72604 standard; Protein; 479 AA.  
 AC R72604:  
 DT 01-DEC-1995 (first entry)  
 DE Human calcium channel subunit beta 1-1.  
 KW Calcium channel subunit; antagonist; agonist; diagnosis;  
 KW Lambert Eaton Syndrome.  
 OS Homo sapiens.  
 PN W09504822-A.  
 PD 16-FEB-1995.  
 PF 11-AUG-1994; U09230.  
 PR 11-AUG-1993; US-105536.  
 PR 05-NOV-1993; US-149097.  
 PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.  
 PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;  
 DR WPI: 95-090900/12.  
 DR N-PSDB: Q87831.  
 PT DNA encoding human calcium channel sub-unit(s) - used for  
 PT developing prods. for studying calcium channels, e.g. for  
 PT obtaining agonists and antagonists  
 PS Disclosure: Page 253-256; 285pp; English.  
 CC To isolate DNA encoding the Beta 1 subunit, a human hippocampus  
 CC encoding a rabbit skeletal muscle calcium channel beta subunit.  
 CC A hybridising clone was selected and was in turn used to isolate  
 CC overlapping clones until the overlapping clones encompassing DNA  
 CC encoding the entire human calcium channel beta 2 subunit were  
 CC isolated and sequenced. Five alternatively spliced forms of the  
 CC beta 1 subunit have been identified. These forms are designated  
 CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in the  
 CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in  
 CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293  
 CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3  
 CC subunits have been constructed. The subunits beta 1-1, beta 1-2,  
 CC beta 1-4 and beta 1-5 have been identified by nucleic acid  
 CC amplification analysis as alternatively spliced forms of the  
 CC beta subunit. Sequences of the beta 1 splice variants are set  
 CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and  
 CC R72604-R72706.  
 SQ Sequence 479 AA;  
 Query Match 79.4%; Score 54; DB 14; Length 479;  
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 44 stsdtsnsfv 55  
 QY 81 STSNDTTSAAVF 92  
 III:IIII:II  
 RESULT 11  
 ID W63156 standard; Protein; 523 AA.  
 AC W63156:  
 DT 12-OCT-1998 (first entry)  
 DE Human calcium channel beta-1 subunit.

KW Beta-1 subunit; human; calcium channel; assay; detection;  
 KW Characterisation; Lambert Eaton Syndrome; LES; diagnosis.  
 OS Homo sapiens.  
 PN US5792846-A.  
 PD 11-AUG-1998. 455543.  
 PF 31-MAY-1995; US-223305.  
 PR 04-APR-1994; US-223305.  
 PR 04-APR-1988; US-176899.  
 PR 04-APR-1989; US-603751.  
 PR 04-APR-1989; WO-U01408.  
 PR 20-FEB-1990; US-482384.  
 PR 30-NOV-1990; US-620250.  
 PR 15-AUG-1991; US-745206.  
 PR 31-MAY-1995; US-455543.  
 PA (SIBI-) SIBIA NEUROSCIENCES INC.  
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,  
 PI Williams ME;  
 DR WPI: 98-456192/39.  
 DR N-PSDB: V42705.  
 PT DNA encoding human calcium channel alpha 1B subunit protein -  
 PT useful for recombinant production of the channel for screening of  
 PT its modulators, and diagnosis of Lambert Eaton Syndrome  
 PS Claim 3: Columns 223-226; 166pp; English.  
 CC The present sequence represents the beta-1 subunit of a human calcium  
 CC channel. Calcium channels are membrane-spanning, multi-subunit proteins  
 CC that allow controlled entry of calcium ions into cells. This leads  
 CC to depolarisation events required for muscle contraction. The recombinant  
 CC subunit, when expressed with nucleic acids encoding the complete calcium  
 CC channel, can be used in assays for the detection and characterisation of  
 CC compounds that modulate the channel. The DNA encoding the subunits can  
 CC be alternatively spliced when transcribed, giving more than one form of  
 CC the protein from the same transcript, each having slightly different  
 CC properties. In addition, the reactivity of the alpha 1 subunit with IgG  
 CC molecules from the serum of an individual with Lambert Eaton Syndrome  
 CC (LES) can be used as a diagnostic for the disease.  
 SQ Sequence 523 AA;

Query Match 79.4%; Score 54; DB 33; Length 523;  
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnf 55  
 |||:||||: ||  
 QY 81 STSNDTTSAAVF 92

RESULT 12  
 ID W63147 standard; Protein: 530 AA;  
 AC W63147;  
 DT 12-OCT-1998 (first entry)  
 DE Human calcium channel beta subunit.  
 KW Beta subunit; human; calcium channel; assay; detection;  
 KW Characterisation; Lambert Eaton Syndrome; LES; diagnosis.  
 OS Homo sapiens.  
 PN US5792846-A.  
 PD 11-AUG-1998. 455543.  
 PF 31-MAY-1995; US-223305.  
 PR 04-APR-1994; US-223305.  
 PR 04-APR-1988; US-176899.  
 PR 04-APR-1989; US-603751.  
 PR 04-APR-1989; WO-U01408.  
 PR 20-FEB-1990; US-482384.  
 PR 30-NOV-1990; US-620250.  
 PR 15-AUG-1991; US-745206.  
 PR 31-MAY-1995; US-455543.  
 PA (SIBI-) SIBIA NEUROSCIENCES INC.  
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,  
 PI Williams ME;  
 DR WPI: 98-456192/39.  
 DR N-PSDB: V42693.  
 PT DNA encoding human calcium channel alpha 1B subunit protein -  
 PT useful for recombinant production of the channel for screening of  
 PT its modulators, and diagnosis of Lambert Eaton Syndrome

PS Disclosure; Columns 125-130; 166pp; English.  
 CC The present sequence represents the beta subunit of a human calcium  
 CC channel. Calcium channels are membrane-spanning, multi-subunit  
 CC proteins that allow controlled entry of calcium ions into cells.  
 CC This leads to depolarisation events required for muscle contraction.  
 CC The recombinant subunit, when expressed with nucleic acids encoding  
 CC the complete calcium channel, can be used in assays for the detection  
 CC and characterisation of compounds that modulate the channel. The  
 CC DNA encoding the subunits can be alternatively spliced when  
 CC transcribed, giving more than one form of the protein from the same  
 CC transcript, each having slightly different properties. In addition, the  
 CC reactivity of the alpha 1 subunit with IgG molecules from the serum of  
 CC an individual with Lambert Eaton Syndrome (LES) can be used as a  
 CC diagnostic for the disease.  
 SQ Sequence 530 AA;

Query Match 79.4%; Score 54; DB 33; Length 530;  
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnf 55  
 |||:||||: ||  
 QY 81 STSNDTTSAAVF 92

RESULT 13  
 ID R39565 standard; Protein: 571 AA.  
 AC R39565;  
 DT 09-FEB-1994 (first entry)  
 DE Human neuronal VDCC beta-subunit encoded by clone HBB3.  
 KW Voltage-dependent calcium channel; VDCC; beta-subunit;  
 KW calcium flux; ss.  
 OS Homo sapiens.  
 PN DE422126-A.  
 PD 19-AUG-1993. 222126.  
 PF 06-JUL-1992; 222126.  
 PR 17-FEB-1992; DE-204716.  
 PR 06-JUL-1992; DE-222126.  
 PA (FARB ) BAYER AG.  
 PI Spreyer P, Unterbeck A;  
 DR WPI: 93-265734/34.  
 DR N-PSDB: Q46078.  
 PT Human neuronal beta-unit cDNA of voltage dependent calcium  
 PT channels - useful in calcium-flux studies and screening systems  
 PT for agonists and antagonists of calcium channels  
 PS Claim 2: Page 11-13; 13pp; German.  
 CC A first oligonucleotide probe (Q46075) complementary to nucleotides  
 CC 361-400 of the VDCC beta-subunit from rabbit skeletal muscle was  
 CC used to screen a human hippocampus cDNA library. A 1.9kb cDNA  
 CC fragment was isolated for further screening to isolate human VDCC  
 CC beta-subunit coding sequences. Clone HBB3 was sequenced and  
 CC nucleotides 1-1288 were found to have 92% homology to the rabbit  
 CC beta-subunit; from position 1289 there is no detectable homology.  
 CC The amino acid sequence R39565 was deduced from the open reading  
 CC frame. See also Q46076 and Q46077.  
 SQ Sequence 571 AA;

Query Match 79.4%; Score 54; DB 8; Length 571;  
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnf 55  
 |||:||||: ||  
 QY 81 STSNDTTSAAVF 92

RESULT 14  
 ID W63144 standard; Protein: 598 AA.  
 AC W63144;  
 DT 12-OCT-1998 (first entry)  
 DE Human calcium channel beta subunit splice variant beta3.  
 KW Beta subunit; human; calcium channel; assay; detection;  
 KW Characterisation; Lambert Eaton Syndrome; LES; diagnosis.

OS Homo sapiens.  
PN US5792846-A.  
PD 11-AUG-1998.  
PF 31-MAY-1995; 455543.  
PR 04-APR-1994; US-223305.  
PR 04-APR-1988; US-176899.  
PR 04-APR-1989; US-603753.  
PR 04-APR-1989; WO-U01408.  
PR 20-FEB-1990; US-482384.  
PR 30-NOV-1990; US-620250.  
PR 15-AUG-1991; US-745206.  
PR 31-MAY-1995; US-455543.  
PA (SIBI-) SIBIA NEUROSCIENCES INC.  
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,  
PI Williams ME;  
DR WPI: 98-456192/39.  
DR N-PSDB; V42688.  
DR DNA encoding human calcium channel alpha 1B subunit protein -  
PT useful for recombinant production of the channel for screening of  
PT its modulators, and diagnosis of Lambert Eaton Syndrome  
PS Claim 3: Columns 219-224; 166pp; English.  
CC The present sequence represents a splice variant of the beta subunit of  
CC a human calcium channel. Calcium channels are membrane-spanning,  
CC multi-subunit proteins that allow controlled entry of calcium ions into  
CC cells. This leads to depolarisation events required for muscle  
CC contraction. The recombinant subunit, when expressed with nucleic acids  
CC encoding the complete calcium channel, can be used in assays for the  
CC detection and characterisation of compounds that modulate the channel.  
CC The DNA encoding the subunits can be alternatively spliced when  
CC transcribed, giving more than one form of the protein from the same  
CC transcript, each having slightly different properties. In addition, the  
CC reactivity of the alpha 1 subunit with IgG molecules from the serum of  
CC an individual with Lambert Eaton Syndrome (LES) can be used as a  
CC diagnostic for the disease.  
SQ Sequence 598 AA;

Query Match 79.4%; Score 54; DB 33; Length 598;  
Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfv 55  
|||:||||:|  
QY 81 STSNDTTSAAVF 92

Search completed: Wed Sep 1 16:33:43 1999  
Job time : 36 secs.

CC beta 1 subunit have been identified. These forms are designated  
CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in the  
CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in  
CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293  
CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3  
CC subunits have been constructed. The subunits beta 1-1, beta 1-2,  
CC beta 1-4 and beta 1-5 have been identified by nucleic acid  
CC amplification analysis as alternatively spliced forms of the  
CC beta subunit. Sequences of the beta 1 splice variants are set  
CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and  
CC R72604-R72706.  
SQ Sequence 598 AA;

Query Match 79.4%; Score 54; DB 14; Length 598;  
Best Local Similarity 75.0%; Pred. No. 1.78e+01;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfv 55  
|||:||||:|  
QY 81 STSNDTTSAAVF 92

Search completed: Wed Sep 1 16:33:43 1999  
Job time : 36 secs.

RESULT 15  
ID R72612 standard; Protein; 598 AA.  
AC R72612;  
DT 01-DEC-1995 (first entry)  
DE Human calcium channel subunit beta 1-3.  
KW Calcium channel subunit; antagonist; agonist; diagnosis;  
KW Lambert Eaton Syndrome.  
OS Homo sapiens.  
PN W09504822-A.  
PD 16-FEB-1995.  
PF 11-AUG-1994; U09230.  
PR 11-AUG-1993; US-105536.  
PR 05-NOV-1993; US-149097.  
PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.  
PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;  
DR WPI: 95-090900/12.  
DR N-PSDB; Q87839.  
PT DNA encoding human calcium channel sub-unit(s) - used for  
PT developing prods. for studying calcium channels, e.g. for  
PT obtaining agonists and antagonists  
PS Disclosure; Page 163-165; 285pp; English.  
CC To isolate DNA encoding the Beta 1 subunit, a human hippocampus  
CC cDNA library was screened by hybridisation to a DNA fragment  
CC encoding a rabbit skeletal muscle calcium channel beta subunit.  
CC A hybridising clone was selected and was in turn used to isolate  
CC overlapping clones until the overlapping clones encompassing DNA  
CC encoding the entire human calcium channel beta 2 subunit were  
CC isolated and sequenced. Five alternatively spliced forms of the

\*\*\*\*\*

W P E L H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:32:35 1999; MasPar time 4.57 Seconds  
Tabular output not generated. 113.947 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (81-93) from PCTUS9913024.pep (12 of 12)  
Perfect Score: 68  
Sequence: 1 STSNDTTSAAAFVS 13

Scoring table: PAM 150  
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 22.122; Variance 24.813; scale 0.892

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	54	79.4	475	2	A44461 voltage-dependent cal	1.44e-01
2	54	79.4	477	2	I65766 L-type voltage-gated	1.44e-01
3	54	79.4	478	2	JH0566 calcium channel beta-	1.44e-01
4	54	79.4	478	2	B44461 voltage-dependent cal	1.44e-01
5	54	79.4	522	2	I65767 L-type voltage-gated	1.44e-01
6	54	79.4	524	2	A41347 calcium channel prote	1.44e-01
7	54	79.4	567	2	A48895 myasthenic syndrome a	1.44e-01
8	54	79.4	596	2	C44461 voltage-dependent cal	1.44e-01
9	54	79.4	597	2	S18304 calcium channel prote	1.44e-01
10	54	79.4	597	2	I52859 L-type voltage-gated	1.44e-01
11	54	79.4	632	2	S21048 calcium channel prote	1.44e-01
12	48	70.6	355	2	D56271 long polar fibrillar o	3.61e+00
13	48	70.6	886	2	A57172 probable hormone rece	3.61e+00
14	48	70.6	1036	1	A34752 nitrogen regulatory p	3.61e+00
15	45	67.6	1076	2	S50536 adenylosuccinate synt	9.97e+00
16	45	66.2	345	2	A64370 hypothetical protei	1.64e+01
17	45	66.2	416	2	S09761 hypothetical protei	1.64e+01
18	45	66.2	583	2	S50959 probable membrane pro	2.67e+01
19	44	64.7	37	2	A21112 variant surface glyco	2.67e+01
20	44	64.7	236	2	I57681 potassium channel pro	2.67e+01
21	44	64.7	490	2	A35312 potassium channel pro	2.67e+01
22	44	64.7	588	2	A39624 probable helicase (EC	2.67e+01
23	44	64.7	1751	1	MMHUMH laminin alpha-2 chain	2.67e+01

24 44 64.7 3106 1 S53868 laminin alpha-2 chain 2.67e-01  
25 43 63.2 110 2 S65003 hypothetical protein 4.31e-01  
26 43 63.2 136 2 B56338 phospholipase A2 (EC 4.31e-01  
27 43 63.2 142 2 A39418 DNA-directed RNA poly 4.31e-01  
28 43 63.2 269 2 S73463 5,10-methylene-tetra 4.31e-01  
29 43 63.2 290 2 B69772 conserved hypothetical 4.31e-01  
30 43 63.2 337 2 S50442 hypothetical protein 4.31e-01  
31 43 63.2 457 2 S60597 1-phosphatidylinosito 4.31e-01  
32 43 63.2 482 2 B44268 cactus - fruit fly (D 4.31e-01  
33 43 63.2 480 2 A39113 potassium channel pro 4.31e-01  
34 43 63.2 500 2 A44268 cactus - fruit fly (D 4.31e-01  
35 43 63.2 500 2 A44268 ankyrin repeat acidic 4.31e-01  
36 43 63.2 548 2 E71961 probable transporter 4.31e-01  
37 43 63.2 549 2 G69618 dipeptide ABC transpo 4.31e-01  
38 43 63.2 552 2 F64546 sodium-dependent tran 4.31e-01  
39 43 63.2 573 2 S50661 hypothetical protein 4.31e-01  
40 43 63.2 630 2 JU0271 voltage-sensitive pot 4.31e-01  
41 43 63.2 757 2 S48841 secretory component p 4.31e-01  
42 43 63.2 757 2 I45956 polymeric immunoglob 4.31e-01  
43 43 63.2 1102 2 S44772 C2954.4 protein - Cae 4.31e-01  
44 43 63.2 1124 2 JH0588 calmodulin-binding pr 4.31e-01  
45 43 63.2 2450 2 S71625 protein-tyrosine-phos 4.31e-01

#### ALIGNMENTS

RESULT 1  
ENTRY #type complete  
TITLE voltage-dependent calcium channel beta subunit betam isoform  
ORGANISM Homo sapiens #common\_name man  
DATE 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change  
ACCESSION A44461; S31919  
REFERENCE A44461  
#authors Powers, P.A.; Liu, S.; Hogan, K.; Gregg, R.G.  
#journal J. Biol. Chem. (1992) 267:22967-22972  
#title Skeletal muscle and brain isoforms of a beta-subunit of human voltage-dependent calcium channels are encoded by a single gene.  
#cross-references MIM:93054516  
#accession A44461  
#status preliminary  
#molecule\_type nucleic acid  
#residues 1-475 #label POW  
#experimental\_source skeletal muscle  
#note sequence inconsistent with the nucleotide translation  
#note sequence extracted from NCBI backbone (NCBIP:118128)  
REFERENCE S31919  
#authors Iles, D.E.; Segers, B.; Sengers, R.C.; Monsieurs, K.; Heytens, L.; Halsall, J.; Hopkins, P.; Ellis, R.; Hall-Curran, J.; Stewart, A.; Wieringa, B.  
#submission submitted to the EMBL Data Library, February 1993  
#description Genetic mapping of the beta and gamma subunits of the L-type voltage-dependent calcium channel on chromosome 17q and exclusion as candidate genes for malignant hyperthermia susceptibility.  
#accession S31919  
#status preliminary  
#molecule\_type DNA  
#residues 146-183 #label ILE  
#cross-references EMBL:221725; NID:g38562; PID:g38563  
SUMMARY #length 475 #molecular-weight 53079 #checksum 7282

Query Match 79.4%; Score 54; DB 2: Length 475;  
Best Local Similarity 75.0%; Pred. No. 1.44e-01;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 44 STSNDTTSNFV 55

QY 81 STSNDTTSAAVF 92

```
RESULT 2
ENTRY
TITLE I65766 #type complete
ORGANISM L-type voltage-gated calcium channel B subunit - human
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
12-Aug-1996
ACCESSIONS I65766
REFERENCE I52859
#authors Collin, T.; Wang, J.; Nargeot, J.; Schwartz, A.
#journal Circ. Res. (1993) 72:1337-1344
#title Molecular cloning of three isoforms of the L-type
voltage-dependent calcium channel B subunit from normal
human heart.
#cross-references MUID:93265672
#accession I65766
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-477 #label RES
#cross-references GB:L06111; NID:gl87016; PID:gl87017
SUMMARY #length 477 #molecular-weight 52977 #checksum 4957
Query Match 79.4%; Score 54; DB 2; Length 477;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 43 STSDTTNSFV 54
| | | : | | | : | |
QY 81 STSNDTTSAAVF 92
RESULT 3
ENTRY
TITLE JH0566 #type complete
ORGANISM calcium channel beta-2 chain - human
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
10-Sep-1997
ACCESSIONS JH0566
REFERENCE JH0564
#authors Williams, M.E.; Feldman, D.H.; McCue, A.F.; Brenner, R.;
Velicelebi, G.; Ellis, S.B.; Harpold, M.M.
#journal Neuron (1992) 8:71-84
#title Structure and functional expression of alpha1, alpha2, and
beta subunits of a novel human neuronal calcium channel
subtype.
#cross-references MUID:92110010
#accession JH0566
#molecule_type mRNA
#residues 1-478 #label WIL
#cross-references GB:M76560; NID:gl79743; PID:gl79744
#experimental_source hippocampus
COMMENT This protein is a subunit of the voltage dependent calcium channel.
KEYWORDS glycoprotein; phosphoprotein
FEATURE
32,167,209,348,374, #binding_site phosphate (Ser) (covalent) (by protein
450,464 kinase C) #status predicted\
64,201 #binding_site phosphate (Thr) (covalent) (by protein
189,425 kinase C) #status predicted\
#binding_site carbohydrate (Asn) (covalent) #status
predicted\
205 #binding_site phosphate (Thr) (covalent) (by
CAMP-dependent kinase) #status predicted
SUMMARY #length 478 #molecular-weight 52934 #checksum 6465
Query Match 79.4%; Score 54; DB 2; Length 478;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 44 STSDTTNSFV 55
| | | : | | | : | |
QY 81 STSNDTTSAAVF 92
RESULT 4
ENTRY
TITLE B44461 #type complete
ORGANISM voltage-dependent calcium channel beta subunit beta1B1
DATE 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change
20-Mar-1998
ACCESSIONS B44461
REFERENCE A44461
#authors Powers, P.A.; Liu, S.; Hogan, K.; Gregg, R.G.
#journal J. Biol. Chem. (1992) 267:22967-22972
#title Skeletal muscle and brain isoforms of a beta-subunit of human
voltage-dependent calcium channels are encoded by a single
gene.
#cross-references MUID:93054616
#accession B44461
#status preliminary
#molecule_type nucleic acid
#residues 1-478 #label POW
#cross-references GB:M92302; NID:gl79803; PID:gl79804
#experimental_source hippocampus
#note sequence inconsistent with the nucleotide translation
#note sequence extracted from NCBI backbone (NCBIP:118131)
SUMMARY #length 478 #molecular-weight 53176 #checksum 7630
Query Match 79.4%; Score 54; DB 2; Length 478;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 44 STSDTTNSFV 55
| | | : | | | : | |
QY 81 STSNDTTSAAVF 92
RESULT 5
ENTRY
TITLE I65767 #type complete
ORGANISM L-type voltage-gated calcium channel B subunit - human
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
12-Aug-1996
ACCESSIONS I65767
REFERENCE I52859
#authors Collin, T.; Wang, J.; Nargeot, J.; Schwartz, A.
#journal Circ. Res. (1993) 72:1337-1344
#title Molecular cloning of three isoforms of the L-type
voltage-dependent calcium channel B subunit from normal
human heart.
#cross-references MUID:93265672
#accession I65767
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-522 #label RES
#cross-references GB:L06112; NID:gl87018; PID:gl87019
SUMMARY #length 522 #molecular-weight 57834 #checksum 7462
Query Match 79.4%; Score 54; DB 2; Length 522;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 43 STSDTTNSFV 54
| | | : | | | : | |
QY 81 STSNDTTSAAVF 92
RESULT 6
ENTRY
TITLE A41347 #type complete
ORGANISM calcium channel protein beta chain,
dihydropyridine-sensitive, skeletal muscle - rabbit
#formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change
07-Feb-1997
ACCESSIONS A41347
```

```
RESULT 4
ENTRY
TITLE B44461 #type complete
ORGANISM voltage-dependent calcium channel beta subunit beta1B1
DATE 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change
20-Mar-1998
ACCESSIONS B44461
REFERENCE A44461
#authors Powers, P.A.; Liu, S.; Hogan, K.; Gregg, R.G.
#journal J. Biol. Chem. (1992) 267:22967-22972
#title Skeletal muscle and brain isoforms of a beta-subunit of human
voltage-dependent calcium channels are encoded by a single
gene.
#cross-references MUID:93054616
#accession B44461
#status preliminary
#molecule_type nucleic acid
#residues 1-478 #label POW
#cross-references GB:M92302; NID:gl79803; PID:gl79804
#experimental_source hippocampus
#note sequence inconsistent with the nucleotide translation
#note sequence extracted from NCBI backbone (NCBIP:118131)
SUMMARY #length 478 #molecular-weight 53176 #checksum 7630
Query Match 79.4%; Score 54; DB 2; Length 478;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 44 STSDTTNSFV 55
| | | : | | | : | |
QY 81 STSNDTTSAAVF 92
RESULT 5
ENTRY
TITLE I65767 #type complete
ORGANISM L-type voltage-gated calcium channel B subunit - human
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
12-Aug-1996
ACCESSIONS I65767
REFERENCE I52859
#authors Collin, T.; Wang, J.; Nargeot, J.; Schwartz, A.
#journal Circ. Res. (1993) 72:1337-1344
#title Molecular cloning of three isoforms of the L-type
voltage-dependent calcium channel B subunit from normal
human heart.
#cross-references MUID:93265672
#accession I65767
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-522 #label RES
#cross-references GB:L06112; NID:gl87018; PID:gl87019
SUMMARY #length 522 #molecular-weight 57834 #checksum 7462
Query Match 79.4%; Score 54; DB 2; Length 522;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 43 STSDTTNSFV 54
| | | : | | | : | |
QY 81 STSNDTTSAAVF 92
RESULT 6
ENTRY
TITLE A41347 #type complete
ORGANISM calcium channel protein beta chain,
dihydropyridine-sensitive, skeletal muscle - rabbit
#formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change
07-Feb-1997
ACCESSIONS A41347
```

```

REFERENCE      A41347
#authors      Ruth, P.; Roehrkasten, A.; Biel, M.; Bosse, E.; Regulla, S.;
#journal      Meyer, H.E.; Flockerzi, V.; Hofmann, F.
#title       Science (1989) 245:1115-1118
#cross-references GB:M25817
#accession   A41347
#status      preliminary
#molecule_type mRNA
#residues    1-524 #label RUT
#cross-references GB:M25817
#keywords    skeletal muscle
#note        length 524 #molecular-weight 57868 #checksum 8438
SUMMARY
Query Match      79.4%; Score 54; DB 2; Length 524;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTNSFV 55
    |||:||||:|
Qy 81 STSNDTTSAFV 92

RESULT 7
ENTRY   A48895 #type complete
TITLE   myasthenic syndrome antigen B - human
ALTERNATE_NAMES
ORGANISM MYSB
#formal_name Homo sapiens #common_name man
DATE 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change
20-Mar-1998
ACCESSIONS
REFERENCE A48895
#authors Rosenfeld, M.R.; Wong, E.; Dalmau, J.; Manley, G.; Posner,
J.B.; Sher, E.; Furneaux, H.M.
#journal Ann. Neurol. (1993) 33:113-120
#title Cloning and characterization of a Lambert-Eaton myasthenic
syndrome antigen.
#cross-references MUID:93263585
#accession A48895
#status preliminary
#residues 1-567 #label ROS
#cross-references GB:S60415; NID:G300416; PID:G300417
#experimental_source fetal brain
#note sequence extracted from NCBI backbone (NCBIN:132135,
NCBIP:132136)
KEYWORDS alternative splicing
SUMMARY length 567 #molecular-weight 62087 #checksum 9649
Query Match      79.4%; Score 54; DB 2; Length 567;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 58 STSDTTNSFV 69
    |||:||||:|
Qy 81 STSNDTTSAFV 92

RESULT 8
ENTRY   C44461 #type complete
TITLE   voltage-dependent calcium channel beta subunit beta1B2
ORGANISM isoform - human
#formal_name Homo sapiens #common_name man
DATE 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change
20-Mar-1998
ACCESSIONS
REFERENCE C44461
#authors Powers, P.A.; Liu, S.; Hogan, K.; Gregg, R.G.
#journal J. Biol. Chem. (1992) 267:22967-22972
#title Skeletal muscle and brain isoforms of a beta-subunit of human
voltage-dependent calcium channels are encoded by a single
gene.

```

```

#cross-references MUID:93054616
#accession C44461
#status preliminary; not compared with conceptual translation
#molecule_type nucleic acid
#residues 1-596 #label POW
#cross-references GB:M92303; NID:G291880; PID:G179806
#experimental_source hippocampus
#note sequence extracted from NCBI backbone (NCBIP:118133)
SUMMARY length 596 #molecular-weight 65808 #checksum 4540
Query Match      79.4%; Score 54; DB 2; Length 596;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTNSFV 55
    |||:||||:|
Qy 81 STSNDTTSAFV 92

RESULT 9
ENTRY   S18304 #type complete
TITLE   calcium channel protein beta chain - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
20-Mar-1998
ACCESSIONS
REFERENCE S18304
#authors Pragnell, M.; Sakamoto, J.; Jay, S.D.; Campbell, K.P.
#journal FEBS Lett. (1991) 291:253-258
#title Cloning and tissue-specific expression of the brain calcium
channel beta-subunit.
#cross-references MUID:92038046
#accession S18304
#molecule_type mRNA
#residues 1-597 #label PRA
#cross-references GB:X61394; NID:G55893; PID:G55894
SUMMARY length 597 #molecular-weight 65679 #checksum 8572
Query Match      79.4%; Score 54; DB 2; Length 597;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTNSFV 55
    |||:||||:|
Qy 81 STSNDTTSAFV 92

RESULT 10
ENTRY   I52859 #type complete
TITLE   L-type voltage-gated calcium channel B subunit - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
02-Jul-1996
ACCESSIONS
REFERENCE I52859
#authors Collin, T.; Wang, J.; Nargeot, J.; Schwartz, A.
#journal Circ. Res. (1993) 72:1337-1344
#title Molecular cloning of three isoforms of the L-type
voltage-dependent calcium channel B subunit from normal
human heart.
#cross-references MUID:93265672
#accession I52859
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-597 #label RES
#cross-references GB:L06110; NID:G187014; PID:G187015
SUMMARY length 597 #molecular-weight 65578 #checksum 5631
Query Match      79.4%; Score 54; DB 2; Length 597;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTNSFV 55

```

```

||||:||||:|
QY 81 STSNDTTSAAVF 92

RESULT 11
ENTRY S21048 #type complete
TITLE calcium channel protein beta chain CaB2b - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
DATE 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change
10-Sep-1997

ACCESSIONS S21048
REFERENCE S21046
#authors Hullin, R.; Singer-Lahat, D.; Freichel, M.; Biel, M.; Dascal,
N.; Hofmann, F.; Flockerzi, V.
#journal EMBO J. (1992) 11:885-890
#title Calcium channel beta subunit heterogeneity: functional
expression of cloned cDNA from heart, aorta and brain.
#cross-references MUID:92192022
#accession S21048
#molecule_type mRNA
#residues 1-632 #label HUL
#cross-references EMBL:X64298; NID:g1499; PID:g1500
SUMMARY #length 632 #molecular-weight 70943 #checksum 1583

Query Match 79.4%; Score 54; DB 2; Length 632;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 30 STSSDTSNSFV 41
||||:||||:|
QY 81 STSNDTTSAAVF 92

RESULT 12
ENTRY D56271 #type complete
TITLE long polar fibrillar operon protein LpfD - Salmonella
ORGANISM typhimurium
DATE 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change
09-Sep-1997

ACCESSIONS D56271
REFERENCE A56271
#authors Baemler, A.J.; Heffron, F.
#journal J. Bacteriol. (1995) 177:2087-2097
#title Identification and sequence analysis of lpfABCD, a putative
fibrillar operon of Salmonella typhimurium.
#accession D56271
#status preliminary
#molecule_type DNA
#residues 1-355 #label BAE
#cross-references GB:U18559; NID:g829370; PID:g829374
GENETICS lpfD
#gene
SUMMARY #length 355 #molecular-weight 37714 #checksum 5448

Query Match 70.6%; Score 48; DB 2; Length 355;
Best Local Similarity 63.6%; Pred. No. 3.61e+00;
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 51 TSKNTTGATFV 61
||||:||||:|
QY 82 TSNDTTSAAVF 92

RESULT 13
ENTRY A57172 #type complete
TITLE probable hormone receptor EMRI precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change
24-Sep-1998
ACCESSIONS A57172
REFERENCE A57172
#authors

```

---

```

#authors Baud, V.; Chisoe, S.L.; Viegas-Pequignot, E.; Diriong, S.;
N'Guyen, V.C.; Roe, B.A.; Lipinski, M.
#journal Genomics (1995) 26:334-344
#title EMRI, an unusual member in the family of hormone receptors
with seven transmembrane segments.
#accession A57172
#status preliminary
#molecule_type mRNA
#residues 1-886 #label BAU
#cross-references GB:X81479; NID:g784993; PID:g784994
GENETICS
#gene GDB:EMRI
#cross-references GDB:378349; OMIM:600493
#map_position 19p13.3-19p13.3
KEYWORDS transmembrane protein
SUMMARY #length 886 #molecular-weight 97679 #checksum 2055

Query Match 70.6%; Score 48; DB 2; Length 886;
Best Local Similarity 61.5%; Pred. No. 3.61e+00;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 470 SESTETGVAFVS 482
||||:||||:|
QY 81 STSNDTTSAAVF 93

RESULT 14
ENTRY A34755 #type complete
TITLE nitrogen regulatory protein nit-2 - Neurospora crassa
ORGANISM #formal_name Neurospora crassa
DATE 13-Jul-1990 #sequence_revision 26-Jul-1996 #text_change
16-Feb-1997

ACCESSIONS A34755
REFERENCE A34755
#authors Fu, Y.H.; Marzluf, G.A.
#journal Mol. Cell. Biol. (1990) 10:1056-1065
#title nit-2, the major nitrogen regulatory gene of Neurospora
crassa, encodes a protein with a putative zinc finger
DNA-binding domain.
#cross-references MUID:90158568
#accession A34755
#molecule_type DNA; mRNA
#residues 1-1036 #label FUY
#cross-references GB:M33956
GENETICS 209/2; 335/3
#introns
CLASSIFICATION #superfamily nitrogen regulatory protein nit-2; GATA-type
zinc finger homology
KEYWORDS DNA binding; transcription regulation; zinc finger
FEATURE 740-793
743-767, #domain GATA-type zinc finger homology #label GZF\
SUMMARY #length 1036 #molecular-weight 109295 #checksum 235

Query Match 70.6%; Score 48; DB 1; Length 1036;
Best Local Similarity 53.8%; Pred. No. 3.61e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 87 SSSNNNTSNGFVA 99
||||:||||:|
QY 81 STSNDTTSAAVF 93

RESULT 15
ENTRY S50536 #type complete
TITLE hypothetical protein YER033c - yeast (Saccharomyces
cerevisiae)
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 28-May-1993 #sequence_revision 24-Feb-1995 #text_change
21-Nov-1997
ACCESSIONS S50536
REFERENCE S50428
#authors Dietrich, F.S.

```

#submission submitted to the EMBL Data Library, December 1994  
#description The sequence of S. cerevisiae cosmids 9379, 9581, and lambda  
clone 4678.

#accession S50536

##molecule\_type DNA

##residues\_type 1-1076 ##label DIE

##cross-references EMBL:U18796; NID:g603265; PID:g603266; MIPS:YER033c

# GENETICS

#map\_position 5R

SUMMARY #length 1076 #molecular-weight 119349 #checksum 9442

Query Match

Best Local Similarity 67.6%; Score 46; DB 2; Length 1076;

Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 97 TSTNOTTSNSFV 108

QY 81 STSNDTTSAAFV 92

Search completed: Wed Sep 1 16:32:49 1999  
Job time : 14 secs.



\*\*\*\*\*  
 WIREH (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Wed Sep 1 16:31:14 1999; MasPar time 2.98 Seconds  
 Tabular output not generated. 123.152 Million cell updates/sec

Title: >PCT-US99-13024-2  
 Description: (81-93) from PCTUS9913024.pep (12 of 12)  
 Perfect Score: 68  
 Sequence: 1 STSNDITSAAFVS 13

Scoring table: PAM 150  
 Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: swiss-prot37  
 1:swissprot

Statistics: Mean 22.697; Variance 22.190; scale 1.023

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	54	79.4	478	1	CCBB_HUMAN	DIHYDROPYRIDINE-SENSIT
2	54	79.4	523	1	CCBC_HUMAN	DIHYDROPYRIDINE-SENSIT
3	54	79.4	524	1	CCBC_RABIT	DIHYDROPYRIDINE-SENSIT
4	54	79.4	567	1	CCB2_HUMAN	DIHYDROPYRIDINE-SENSIT
5	54	79.4	596	1	CCBA_HUMAN	DIHYDROPYRIDINE-SENSIT
6	54	79.4	597	1	CCBA_RAT	DIHYDROPYRIDINE-SENSIT
7	54	79.4	632	1	CCB2_RABIT	DIHYDROPYRIDINE-SENSIT
8	48	70.6	355	1	LPFD_SALTY	LPFD PROTEIN PRECURSOR
9	48	70.6	886	1	EMRL_HUMAN	CELL SURFACE GLYCOPROT
10	48	70.6	1036	1	NIT2_NEUCR	NITROGEN CATABOLIC ENZ
11	46	67.6	532	1	CPPL_DROAC	CITOCROME P450 12B1 P
12	46	67.6	1076	1	YEM3_YEAST	HYPOTHETICAL 119.3 KD
13	45	66.2	345	1	PURA_METJA	ADENYLOSUCGINATE SYNTH
14	45	66.2	416	1	IR12_HCMVA	HYPOTHETICAL PROTEIN 1
15	45	66.2	3066	1	POLG_BCMVN	GENOME POLYPROTEIN [CO
16	44	64.7	490	1	C1KL_DROME	VOLTAGE-GATED POTASSIU
17	44	64.7	588	1	PR28_YEAST	PRE-MRNA SPLICING FACT
18	44	64.7	607	1	HRA1_XANCY	HYPERSENSITIVITY RESPO
19	44	64.7	3106	1	LMA2_MOUSE	LAMININ ALPHA-2 CHAIN
20	44	64.7	3110	1	LMA2_HUMAN	LAMININ ALPHA-2 CHAIN
21	43	63.2	142	1	RPC9_YEAST	DNA-DIRECTED RNA POLYM
22	43	63.2	269	1	FOLD_MYCPN	METHYLENETETRAHYDROFOL
23	43	63.2	337	1	YEB7_YEAST	HYPOTHETICAL 38.2 KD P

24 43 63.2 477 1 AMT\_MYCTU PROBABLE AMMONIUM TRAN 1.75e+01  
 25 43 63.2 500 1 CACT\_DROME DEVELOPMENTAL PROTEIN 1.75e+01  
 26 43 63.2 573 1 YEB8\_YEAST HYPOTHETICAL 63.7 KD P 1.75e+01  
 27 43 63.2 757 1 PIGR\_BOVIN POLYMERIC-IMMUNOGLOBUL 1.75e+01  
 28 43 63.2 786 1 XS8A\_CAEEL HYPOTHETICAL 84.3 KD P 1.75e+01  
 29 43 63.2 1102 1 YK64\_CAEEL HYPOTHETICAL 124.8 KD 1.75e+01  
 30 43 63.2 1124 1 TRPL\_DROME TRANSIENT-RECEPTOR-POT 1.75e+01  
 31 43 63.2 1234 1 JMJ\_MOUSE JUMONJI PROTEIN. 1.75e+01  
 32 42 61.8 290 1 YEAB\_BACSU HYPOTHETICAL 31.8 KD P 2.94e+01  
 33 42 61.8 309 1 YF44\_METH HYPOTHETICAL SUGAR KIN 2.94e+01  
 34 42 61.8 325 1 YHU3\_YEAST HYPOTHETICAL 33.4 KD P 2.94e+01  
 35 42 61.8 331 1 PERX\_SOLTU SUBERIZATION-ASSOCIATE 2.94e+01  
 36 42 61.8 363 1 PER2\_LYCES SUBERIZATION-ASSOCIATE 2.94e+01  
 37 42 61.8 364 1 PER1\_LYCES SUBERIZATION-ASSOCIATE 2.94e+01  
 38 42 61.8 429 1 HISX\_METJA HISTIDINOL DEHYDROGENA 2.94e+01  
 39 42 61.8 457 1 TBB1\_PORPU TUBULIN BETA CHAIN 1. 2.94e+01  
 40 42 61.8 514 1 AMT2\_LYCES HIGH AFFINITY AMMONIUM 2.94e+01  
 41 42 61.8 871 1 YAY7\_SCHPO HYPOTHETICAL 98.4 KD P 2.94e+01  
 42 42 61.8 887 1 YAY5\_SCHPO PROBABLE ATP-DEPENDENT 2.94e+01  
 43 42 61.8 953 1 PSD1\_HUMAN 26S PROTEASOME REGULAT 2.94e+01  
 44 42 61.8 1333 1 YN99\_YEAST PROBABLE ATP-DEPENDENT 2.94e+01  
 45 42 61.8 1475 1 N153\_HUMAN NUCLEAR PORE COMPLEX P 2.94e+01

## ALIGNMENTS

RESULT 1  
 ID CCBB\_HUMAN STANDARD; PRT; 478 AA.  
 AC Q02640;  
 DT 01-JUN-1994 (REL. 29, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-1-B1 SUBUNIT  
 DE (BETA-1 ISOFORM B) (BETA-2) (BETA-1C).  
 GN CACNB1 OR CACNLB1.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-HIPPOCAMPUS;  
 RX MEDLINE; 93054616.  
 RA POWERS P.A., LIU S., HOGAN K., GREGG R.G.;  
 RT "Skeletal muscle and brain isoforms of a beta-subunit of human  
 voltage-dependent calcium channels are encoded by a single gene.";  
 RL J. BIOL. CHEM. 267:22967-22972(1992).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92110010.  
 RA WILLIAMS M.E., FELDMAN D.H., MCCUE A.F., BRENNER R.,  
 RA VELICELEBI G., ELLIS S.B., HARPOLD M.M.;  
 RT "Structure and functional expression of alpha 1, alpha 2, and beta  
 subunits of a novel human neuronal calcium channel subtype.";  
 RL NEURON 8:71-84(1992).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-HEART;  
 RX MEDLINE; 93265672.  
 RA COLLIN T., WANG J., NARGEOT J., SCHWARTZ A.;  
 RT "Molecular cloning of three isoforms of the L-type voltage-dependent  
 calcium channel beta subunit from normal human heart.";  
 RL CIRC. RES. 72:1337-1344(1993).  
 [4]  
 RP SEQUENCE FROM N.A.  
 RA POWERS P.A., GREGG R.G., HOGAN K.;  
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -!- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)  
 CC SENSITIVE CALCIUM CHANNEL.  
 CC -!- SURUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:  
 CC ALPHA-1, ALPHA-2, BETA AND GAMMA.  
 CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH  
 CC THE CYTOPLASMIC ASPECT OF THE SARCOLEMMMA.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, HEART AND SPLEEN.

CC -1- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1  
 CC SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -1- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M92302; G179804; -  
 DR EMBL: M76560; G179744; -  
 DR EMBL: L06111; G187017; -  
 DR EMBL: U86960; G2155254; JOINED.  
 DR EMBL: U86952; G2155254; JOINED.  
 DR EMBL: U86953; G2155254; JOINED.  
 DR EMBL: U86954; G2155254; JOINED.  
 DR EMBL: U86955; G2155254; JOINED.  
 DR EMBL: U86956; G2155254; JOINED.  
 DR EMBL: U86957; G2155254; JOINED.  
 DR EMBL: U86958; G2155254; JOINED.  
 DR EMBL: U86959; G2155254; JOINED.  
 DR PIR: B4461; B4461.  
 DR MIM: I14207; -  
 DR PROSITE: PS50002; SH3; 1.  
 DR PFAM: PF00018; SH3; 1.  
 DR PFAM: PF00774; Ca\_channel\_B; 1.  
 KW IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;  
 KW GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;  
 KW MULTIGENE FAMILY.  
 FT DOMAIN 100 161 SH3.  
 FT CARBOHYD 189 189 POTENTIAL.  
 FT CARBOHYD 425 425 POTENTIAL.  
 FT CONFLICT 21 21 E -> G (IN REF. 2).  
 FT CONFLICT 28 29 QG -> R (IN REF. 3).  
 FT CONFLICT 434 435 AA -> RR (IN REF. 1 AND 4).  
 SQ SEQUENCE 478 AA; 53006 MW; 6011892D CRC32;

Query Match 79.48; Score 54; DB 1; Length 478;

Best Local Similarity 75.08; Pred. No. 3.03e-02;

Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNSFV 55

QY 81 STSDTTSRAFV 92

RESULT 2  
 ID CCBC\_HUMAN STANDARD; PRT; 523 AA.  
 AC Q02639;  
 DT 01-JUN-1994 (REL. 29, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-1M SUBUNIT  
 DE (BETA-1 ISOFORM C) (BETA-1A).  
 GN CACNB1 OR CACNLB1.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA.  
 OC CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-SKELETAL MUSCLE;  
 RX MEDLINE: 93054616.  
 RA POWERS P.A., LIU S., HOGAN K., GREGG R.G.;  
 RT "Skeletal muscle and brain isoforms of a beta-subunit of human  
 RT voltage-dependent calcium channels are encoded by a single gene."  
 RL J. BIOL. CHEM. 267:22967-22972(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-HEART;

RX MEDLINE: 93265672.  
 RA COLLIN T., WANG J., NARGEOT J., SCHWARTZ A.;  
 RT "Molecular cloning of three isoforms of the L-type voltage-dependent  
 RT calcium channel beta subunit from normal human heart."  
 RL CIRC. RES. 72:1337-1344(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA POWERS P.A., GREGG R.G., HOGAN K.;  
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [4]  
 RP SEQUENCE OF 146-209 FROM N.A.  
 RX MEDLINE: 93372845.  
 RA ILES D.E., SEGERS B., SENGERS R.C.A., MONSIEURS K., HEYTENS L.,  
 RA HALSALL P.J., HOPKINS P.M., ELLIS F.R., HALL-CURRAN J.L.,  
 RA STEWART A.D., WIERINGA B.;  
 RT "Genetic mapping of the beta 1- and gamma-subunits of the human  
 RT skeletal muscle L-type voltage-dependent calcium channel on  
 RT chromosome 17q and exclusion as candidate genes for malignant  
 RT hyperthermia susceptibility."  
 RL HUM. MOL. GENET. 2:863-868(1993).  
 CC -1- FUNCTION: THIS PROTEIN IS THE BETA SUBUNIT OF THE DIHYDROPYRIDINE  
 CC (DHP) SENSITIVE CALCIUM CHANNEL. PLAYS A ROLE IN EXCITATION-  
 CC CONTRACTION COUPLING. THE SKELETAL MUSCLE DHP-SENSITIVE CA(2+)-  
 CC CHANNEL MAY FUNCTION ONLY AS A MULTIPLE SUBUNIT COMPLEX.  
 CC -1- SUBUNIT: THE SKELETAL MUSCLE L-TYPE CALCIUM CHANNEL IS COMPOSED OF  
 CC FIVE SUBUNITS: ALPHA-1, ALPHA-2, BETA, GAMMA AND DELTA.  
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH  
 CC THE CYTOPLASMIC ASPECT OF THE SARCOLEMMMA.  
 CC -1- TISSUE SPECIFICITY: SKELETAL MUSCLE.  
 CC -1- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1  
 CC SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -1- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M92301; G179802; -  
 DR EMBL: L06112; G187019; -  
 DR EMBL: U86960; G2155255; -  
 DR EMBL: U86952; G2155255; JOINED.  
 DR EMBL: U86953; G2155255; JOINED.  
 DR EMBL: U86954; G2155255; JOINED.  
 DR EMBL: U86955; G2155255; JOINED.  
 DR EMBL: U86956; G2155255; JOINED.  
 DR EMBL: U86957; G2155255; JOINED.  
 DR EMBL: U86958; G2155255; JOINED.  
 DR EMBL: U86959; G2155255; JOINED.  
 DR EMBL: Z21725; G38563; -  
 DR EMBL: Z21726; G38565; -  
 DR MIM: I14207; -  
 DR PROSITE: PS50002; SH3; 1.  
 DR PFAM: PF00018; SH3; 1.  
 DR PFAM: PF00774; Ca\_channel\_B; 1.  
 KW IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;  
 KW GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;  
 KW MULTIGENE FAMILY.  
 FT DOMAIN 100 161 SH3.  
 FT CARBOHYD 189 189 POTENTIAL.  
 FT CARBOHYD 470 470 POTENTIAL.  
 FT CONFLICT 28 29 QG -> R (IN REF. 2).  
 FT CONFLICT 183 183 MISSING (IN REF. 4).  
 FT CONFLICT 479 480 AA -> RR (IN REF. 1 AND 3).  
 SQ SEQUENCE 523 AA; 57863 MW; 32CBEE9D CRC32;

Query Match 79.48; Score 54; DB 1; Length 523;

Best Local Similarity 75.08; Pred. No. 3.03e-02;

Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNFSV 55  
QY 81 STSNDTTSAAFY 92

RESULT 3  
ID CCB2\_HUMAN STANDARD; PRT: 524 AA.  
AC P19517;  
DT 01-FEB-1991 (REL. 17, CREATED)  
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-1M SUBUNIT  
DE (BETA-1 ISOFORM C).  
GN CACNB1 OR CACNLB1.  
OS ORYCTOLAGUS CUNICULUS (RABBIT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE-SKELETAL MUSCLE;  
RA MEDLINE; 89369946.  
RA RUTH P., ROEHRKASTEN A., BIEL M., BOSSE E., REGULLA S., MEYER H.E.,  
RA FLOCKERZI V., HOFFMANN F.,  
RT "Primary structure of the beta subunit of the DHP-sensitive calcium  
channel from skeletal muscle.";  
RL SCIENCE 245:1115-1118(1989).  
CC -!- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)  
SENSITIVE CALCIUM CHANNEL.  
CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:  
ALPHA-1, ALPHA-2, BETA AND GAMMA.  
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH  
THE CYTOPLASMIC ASPECT OF THE SARCOLEMMMA.  
CC -!- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1  
SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE (BY  
SIMILARITY).  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M25817; G164834; -  
CC EMBL: M25514; G598370; -  
CC PROSITE: PS50002; SH3; 1.  
CC PFAM: PF00018; SH3; 1.  
CC PFAM: PF00774; Ca\_channel\_B; 1.  
CC IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;  
CC GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;  
CC MULTIGENE FAMILY.  
CC MOD\_RES 71 ?1 BLOCKED.  
CC DOMAIN 100 161 SH3.  
CC CARBOHYD 189 189 POTENTIAL.  
CC CARBOHYD 470 470 POTENTIAL.  
CC CARBOHYD 499 499 POTENTIAL.  
CC CONFLICT 304 304 L -> R (IN G598370).  
CC SEQUENCE 524 AA; 57825 MW; 8AAC8B67 CRC32;  
Query Match 79.4%; Score 54; DB 1; Length 524;  
Best Local Similarity 75.0%; Pred. No. 3.03e-02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNFSV 55  
QY 81 STSNDTTSAAFY 92

RESULT 4  
ID CCB2\_HUMAN STANDARD; PRT: 567 AA.  
AC Q08289;  
DT 01-OCT-1994 (REL. 30, CREATED)  
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-2 SUBUNIT  
DE (LAMBERT-EATON MYASTHENIC SYNDROME ANTIGEN B) (MYSB).  
GN CACNB2 OR CACNLB2 OR MYSB.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-FETAL BRAIN;  
RA MEDLINE; 93263585.  
RA ROSENFELD M.R., WONG E., DALMAU J., MANLEY G., POSNER J.B.,  
RA SHER E., FURNEAUX H.M.,  
RT "Cloning and characterization of a Lambert-Eaton myasthenic syndrome  
antigen.";  
RL ANN. NEUROL. 33:113-120(1993).  
CC -!- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)  
SENSITIVE CALCIUM CHANNEL (BY SIMILARITY).  
CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:  
ALPHA-1, ALPHA-2, BETA AND GAMMA.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES.  
CC -!- ALTERNATIVE PRODUCTS: THREE FORMS ARE PRODUCED BY ALTERNATIVE  
SPLICING OF THE GENE.  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: S60415; G300417; -  
CC PIR: A48895; A48895.  
CC MIM: 600003; -  
CC PROSITE: PS50002; SH3; 1.  
CC PFAM: PF00018; SH3; 1.  
CC PFAM: PF00774; Ca\_channel\_B; 1.  
CC IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;  
CC GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;  
CC ANTIGEN.  
CC DOMAIN 114 175 SH3.  
CC CARBOHYD 122 122 POTENTIAL.  
CC CARBOHYD 203 203 POTENTIAL.  
CC CARBOHYD 523 523 POTENTIAL.  
CC VARSPLIC 234 268  
CC VARSPLIC 224 268  
CC AIDIDATGLDAEENDIPANHRSPKPSANSVTSPPHSKEKRM  
CC AIDIDATGLDAEENDIPANHRSPKPSANSVTSPPHSKEKRM  
CC FEKK -> AKQKQKS (IN ALTERNATE FORM 2).  
CC FEKK -> AKQKQKS (IN ALTERNATE FORM 2).  
CC SEQUENCE 567 AA; 62087 MW; 4559C166 CRC32;  
Query Match 79.4%; Score 54; DB 1; Length 567;  
Best Local Similarity 75.0%; Pred. No. 3.03e-02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 58 STSDTTSNFSV 69  
QY 81 STSNDTTSAAFY 92

RESULT 5  
ID CCB2\_HUMAN STANDARD; PRT: 596 AA.  
AC Q02641;  
DT 01-JUN-1994 (REL. 29, CREATED)  
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DIHYDROPYRIDINE-SENSITIVE L-TYPE, CHANNEL BETA-1-B2 SUBUNIT  
(BETA-1 ISOFORM A).  
CACNB1 OR CACNLB1.  
HOMO SAPIENS (HUMAN).  
EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
[1]  
SEQUENCE FROM N.A.  
TISSUE-HIPPOCAMPUS;  
MEDLINE; 93054616;  
POWERS P.A., LIU S., HOGAN K., GREGG R.G.;  
"Skeletal muscle and brain isoforms of a beta-subunit of human  
voltage-dependent calcium channels are encoded by a single gene.";  
J. BIOL. CHEM. 267:22967-22972(1992).  
[2]  
SEQUENCE FROM N.A.  
TISSUE-HEART;  
MEDLINE; 93265672.  
COLLIN T., WANG J., NARGEOT J., SCHWARTZ A.;  
"Molecular cloning of three isoforms of the L-type voltage-dependent  
calcium channel beta subunit from normal human heart.";  
CIRC. RES. 72:1337-1344(1993).  
[3]  
- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)  
SENSITIVE CALCIUM CHANNEL.  
- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:  
ALPHA-1, ALPHA-2, BETA AND GAMMA.  
- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH  
THE CYTOPLASMIC ASPECT OF THE SARCOLEMA.  
- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, HEART AND SPLEEN.  
- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1  
SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE.  
- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.  
-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M92303; G179806; -;  
EMBL; L06110; G187015; -;  
PIR; C44461; C44461.  
MIM; 114207; -;  
PROSITE; PS50002; SH3; 1.  
PFAM; PF00018; SH3; 1.  
PFAM; PF00774; Ca\_channel\_B; 1.  
IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;  
GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;  
MULTIGENE FAMILY.  
DOMAIN 100 161 SH3.  
FT CARBOHYD 189 189 POTENTIAL.  
FT CARBOHYD 425 425 POTENTIAL.  
FT CONFLICT 5 6 TS -> SG (IN REF. 2).  
FT CONFLICT 29 29 G -> R (IN REF. 2).  
FT CONFLICT 135 135 H -> D (IN REF. 2).  
FT CONFLICT 175 176 KL -> TV (IN REF. 2).  
FT CONFLICT 182 182 G -> S (IN REF. 2).  
FT CONFLICT 217 217 T -> S (IN REF. 2).  
FT CONFLICT 293 296 SNTR -> LQHT (IN REF. 2).  
FT CONFLICT 344 344 I -> L (IN REF. 2).  
FT CONFLICT 428 428 M -> I (IN REF. 2).  
FT CONFLICT 434 435 AA -> RR (IN REF. 1).  
FT CONFLICT 456 456 E -> D (IN REF. 2).  
FT CONFLICT 465 465 M -> V (IN REF. 2).  
FT CONFLICT 482 482 S -> N (IN REF. 2).  
FT CONFLICT 492 492 R -> W (IN REF. 2).  
FT CONFLICT 515 515 L -> P (IN REF. 2).  
FT CONFLICT 532 532 L -> P (IN REF. 2).  
FT CONFLICT 539 539 A -> GTP (IN REF. 2).  
FT CONFLICT 548 548 MISSING (IN REF. 2).

FT CONFLICT 557 557 L -> M (IN REF. 2).  
FT CONFLICT 571 571 WP -> CA (IN REF. 2).  
FT CONFLICT 591 591 R -> Q (IN REF. 2).  
SQ SEQUENCE 596 AA; 63638 MW; 0CC447D4 CRC32;  
Query Match 79.4%; Score 54; DB 1; Length 596;  
Best Local Similarity 75.0%; Pred. No. 3.03e-02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Db 44 STSDTTSNSFV 55  
|||:||||:|  
Qy 81 STSNDTTSAAFV 92  
|||:||||:|  
RESULT 6  
ID CCB\_A\_RAT STANDARD; PRT; 597 AA.  
AC P34283;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CHANNEL BETA-1-B2 SUBUNIT  
DE (BETA-1 ISOFORM A).  
GN CACNB1 OR CACNLB1.  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
RN [1]  
RC SEQUENCE FROM N.A.  
RP TISSUE-BRAIN;  
RX MEDLINE; 92038046.  
RA PRAGNELL M., SAKAMOTO J., JAY S.D., CAMPBELL K.P.;  
"Cloning and tissue-specific expression of the brain calcium channel  
beta-subunit".  
FEBS LETT. 291:253-258(1991).  
CC - FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)  
SENSITIVE CALCIUM CHANNEL.  
CC - SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:  
ALPHA-1, ALPHA-2, BETA AND GAMMA.  
CC - SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH  
THE CYTOPLASMIC ASPECT OF THE SARCOLEMA.  
CC - ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1  
SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE (BY  
SIMILARITY).  
CC - SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC - SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.  
-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X61394; G55894; -;  
DR PROSITE; PS50002; SH3; 1.  
DR PFAM; PF00018; SH3; 1.  
DR PFAM; PF00774; Ca\_channel\_B; 1.  
IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;  
GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;  
MULTIGENE FAMILY.  
DOMAIN 100 161 SH3.  
FT CARBOHYD 189 189 POTENTIAL.  
FT CARBOHYD 425 425 POTENTIAL.  
SQ SEQUENCE 597 AA; 65679 MW; 23B08C47 CRC32;  
Query Match 79.4%; Score 54; DB 1; Length 597;  
Best Local Similarity 75.0%; Pred. No. 3.03e-02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Db 44 STSDTTSNSFV 55  
|||:||||:|  
Qy 81 STSNDTTSAAFV 92  
|||:||||:|

```

RESULT 7
ID CCB2_RABIT STANDARD; PRT; 632 AA.
AC P54288;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-2 SUBUNIT
DE (CAB2).
GN CACNB2 OR CACNLB2.
OS ORYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92192022.
RA HULLIN R., SINGER-LAHAT D., FREICHEL M., BIEL M., DASCAL N.,
RA HOFMANN F., FLOCKERZI V.;
RT "Calcium channel beta subunit heterogeneity: functional expression of
RT cloned cDNA from heart, aorta and brain.";
RL EMBO J. 11:885-890(1992).
CC -!- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)
CC SENSITIVE CALCIUM CHANNEL.
CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:
CC ALPHA-1, ALPHA-2, BETA AND GAMMA.
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN HEART, AORTA AND
CC BRAIN.
CC -!- ALTERNATIVE PRODUCTS: THREE FORMS OF THE BETA-2 SUBUNIT ARE
CC PRODUCED BY ALTERNATIVE SPLICING OF THE GENE: CAB2A, CAB2B (SHOWN
CC HERE) AND CAB2C.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: X64298; GI500;
DR EMBL; X64297; GI498;
DR EMBL; X64299; GI502;
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00774; Ca_channel_B; 1.
KW IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;
KW GLYCOPROTEIN; PHOSPHORYLATION; SH3 DOMAIN; MULTIGENE FAMILY;
KW ALTERNATIVE SPLICING.
FT DOMAIN 86 147 SH3.
FT CARBOHYD 94 94 POTENTIAL.
FT CARBOHYD 175 175 POTENTIAL.
FT VARSPIC 1 43
FT VARSPIC 197 234
FT VARSPIC 235 240 MISSING (IN CAB2C).
FT VARSPIC 235 240 MPFFKK -> KQKQKS (IN CAB2C).
SQ SEQUENCE 632 AA; 70943 MW; DF97DA5D CRC32;

Query Match 79.4%; Score 54; DB 1; Length 632;
Best Local Similarity 75.0%; Pred. No. 3.03e-02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 30 STSDTTSNSFV 41
QY 81 STSDTTSNAFV 92

RESULT 8
ID LPFD_SALTY STANDARD; PRT; 355 AA.
AC P43663;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE LPFD PROTEIN PRECURSOR.
GN LPFD.
OS SALMONELLA TYPHIMURUM.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 14028;
RX MEDLINE: 95238281.
RA BAUMLER A.J., HEFFRON F.;
RT Identification and sequence analysis of lpfABCD, a putative
RT fimbrial operon of Salmonella typhimurium.;
RL J. BACTERIOL. 177:2087-2097(1995).
CC -!- SIMILARITY: BELONGS TO THE FIMH / LPFD FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: U18559; G829374;
DR STYGENE; SG10560; LPFD.
DR PFAM; PF00419; Fimbrial; 1.
KW FIMBRIA; SIGNAL.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 355 LPFD PROTEIN.
SQ SEQUENCE 355 AA; 37714 MW; 20D529B5 CRC32;

Query Match 70.6%; Score 48; DB 1; Length 355;
Best Local Similarity 63.6%; Pred. No. 1.11e-00;
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 51 TSKNTTGATFV 61
QY 82 TSNDTTSNAFV 92

RESULT 9
ID EMRL_HUMAN STANDARD; PRT; 886 AA.
AC Q14246;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CELL SURFACE GLYCOPROTEIN EMR1 PRECURSOR (EMR1 HORMONE RECEPTOR).
GN EMR1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95324926.
RA BAUD V., CHISOE S.L., VIEGAS-PEQUIGNOT E., DIRIONG S., N'GUEN V.C.,
RA ROE B.A., LIPINSKI M.;
RT "EMR1, an unusual member in the family of hormone receptors with
RT seven transmembrane segments.";
RL GENOMICS 26:334-344(1995).
CC -!- FUNCTION: PROBABLY INVOLVED IN CELLULAR RESPONSE TO A HORMONE OR
CC AN INTERACTION WITH A PROTEIN LIGAND.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: WIDE EXPRESSION; INCREASED LEVELS IN
CC PERIPHERAL BLOOD MONONUCLEAR CELLS.
CC -!- PTM: N- AND O-GLYCOSYLATED; (POSSIBLE).
CC -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
CC CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL; X81479; G784994; -.  
DR MTM; 600493; -.  
DR PROSITE; PS00649; G\_PROTEIN\_RECP\_F2\_1; FALSE\_NEG.  
DR PROSITE; PS00650; G\_PROTEIN\_RECP\_F2\_2; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 6.  
DR PROSITE; PS01186; EGF\_2; 2.  
DR PROSITE; PS01187; EGF\_CA; 5.  
DR PFAM; PF00008; EGF; 3.  
DR HSP; P00736; IAP0.  
DR G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; RECEPTOR; GLYCOPROTEIN;  
KW EGF-LIKE DOMAIN; REPEAT; SIGNAL.  
FT SIGNAL 1 17 POTENTIAL.  
FT CHAIN 18 886 CELL SURFACE GLYCOPROTEIN EMR1.  
FT DOMAIN 18 599 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 600 627 POTENTIAL.  
FT DOMAIN 628 634 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 635 656 POTENTIAL.  
FT DOMAIN 657 666 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 667 690 POTENTIAL.  
FT DOMAIN 691 709 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 710 731 POTENTIAL.  
FT DOMAIN 732 747 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 748 776 POTENTIAL.  
FT DOMAIN 777 794 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 795 814 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 815 829 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 830 852 POTENTIAL.  
FT DOMAIN 853 886 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 887 917 EGF-LIKE 1.  
FT DOMAIN 918 131 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 132 171 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 172 213 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 214 267 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 268 316 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 317 599 SER/THR-RICH.  
FT DISULFID 35 47 BY SIMILARITY.  
FT DISULFID 41 56 BY SIMILARITY.  
FT DISULFID 58 78 BY SIMILARITY.  
FT DISULFID 84 97 BY SIMILARITY.  
FT DISULFID 91 106 BY SIMILARITY.  
FT DISULFID 108 130 BY SIMILARITY.  
FT DISULFID 136 148 BY SIMILARITY.  
FT DISULFID 142 157 BY SIMILARITY.  
FT DISULFID 159 170 BY SIMILARITY.  
FT DISULFID 176 188 BY SIMILARITY.  
FT DISULFID 182 197 BY SIMILARITY.  
FT DISULFID 199 212 BY SIMILARITY.  
FT DISULFID 225 235 BY SIMILARITY.  
FT DISULFID 229 244 BY SIMILARITY.  
FT DISULFID 246 266 BY SIMILARITY.  
FT DISULFID 272 285 BY SIMILARITY.  
FT DISULFID 279 294 BY SIMILARITY.  
FT DISULFID 296 315 BY SIMILARITY.  
FT CARBOHYD 94 94 POTENTIAL.  
FT CARBOHYD 99 99 POTENTIAL.  
FT CARBOHYD 127 127 POTENTIAL.  
FT CARBOHYD 167 167 POTENTIAL.  
FT CARBOHYD 189 189 POTENTIAL.  
FT CARBOHYD 194 194 POTENTIAL.  
FT CARBOHYD 232 232 POTENTIAL.  
FT CARBOHYD 238 258 POTENTIAL.  
FT CARBOHYD 312 312 POTENTIAL.  
FT CARBOHYD 366 366 POTENTIAL.  
FT CARBOHYD 375 375 POTENTIAL.  
FT CARBOHYD 448 448 POTENTIAL.  
FT CARBOHYD 661 661 POTENTIAL.  
SQ SEQUENCE 886 AA; 97680 MW; 36A07647 CRC32;

-----  
Query Match 70.6%; Score 48; DB 1: Length 886;  
Best Local Similarity 61.5%; Pred. No. 1.11e+00;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
-----  
Db 470 SESTETGVAFVS 482  
QY 81 STSNDTTSAAVFS 93  
| | : | | : | | | |  
| | : | | : | | | |  
RESULT 10  
ID NIT2\_NEUCR STANDARD; PRT: 1036 AA.  
AC P19212;  
DT 01-NOV-1990 (REL. 16, CREATED)  
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE NITROGEN CATABOLIC ENZYME REGULATORY PROTEIN.  
GN NIT-2.  
OS NEUROSPORA CRASSA.  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;  
OC SORDARIALES; SORDARIACEAE; NEUROSPORA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=74-OR23-1A;  
RX MEDLINE; 90158568.  
RA FU Y.-H., MARZLUF G.A.;  
RT 'nit-2, the major nitrogen regulatory gene of Neurospora crassa,  
RT encodes a protein with a putative zinc finger DNA-binding domain.';  
RL MOL. CELL. BIOL. 10:1056-1065(1990).  
RN [2]  
RP MUTAGENESIS.  
RX MEDLINE; 91186820  
RA FU Y.-H., MARZLUF G.A.;  
RT 'Site-directed mutagenesis of the 'zinc finger' DNA-binding domain of  
RT the nitrogen-regulatory protein NIT2 of Neurospora.';  
RL MOL. MICROBIOL. 4:1847-1852(1990).  
CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. DURING CONDITIONS  
CC OF NITROGEN LIMITATION IT TURNS ON THE EXPRESSION OF GENES FOR  
CC ENZYMES WHICH ARE REQUIRED FOR THE USE OF A VARIETY OF SECONDARY  
CC NITROGEN SOURCES, INCLUDING NITRATES, PURINES, AMINO ACIDS, AND  
CC PROTEINS.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- INDUCTION: BY LACK OF A PRIMARY NITROGEN SOURCE.  
CC -!- SIMILARITY: HIGH TO OTHER FUNGAL NITROGEN REGULATORY PROTEINS.  
CC -!- SIMILARITY: TO GATA TRANSCRIPTION FACTORS IN THE ZINC-FINGER  
CC REGION.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL; M33956; G1107471; -.  
DR PIR; A34755; A34755  
DR PROSITE; PS00344; GATA\_ZN\_FINGER; 1.  
DR PFAM; PF00320; GATA; 1.  
DR HSP; P17429; 5GAT.  
DR TRANSFAC; T00627; -.  
KW TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING; ZINC-FINGER;  
KW NUCLEAR PROTEIN; NITRATE ASSIMILATION; REPEAT.  
FT DOMAIN 49 110 3 X APPROXIMATE REPEATS.  
FT REPEAT 49 55 1.  
FT REPEAT 87 92 2.  
FT REPEAT 105 110 3.  
FT ZN\_FING 743 767 GATA-TYPE.  
FT MUTAGEN 743 746 CTNC->STNG; ABOLISH DNA-BINDING.  
FT MUTAGEN 755 756 RR->GG; ABOLISH DNA-BINDING.  
FT MUTAGEN 765 766 NA->DV; ABOLISH DNA-BINDING.  
FT MUTAGEN 768 769 GL->DV; ABOLISH DNA-BINDING.  
FT MUTAGEN 789 790 KR->NS; ABOLISH DNA-BINDING.

\*\*\*\*\*

# WQELH

\*\*\*\*\* (TM) \*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 14:06:15 1999; MasPar time 1940.09 Seconds  
Tabular output not generated. 1351.480 Million cell updates/sec

Title: &gt;PCT-US99-13024-1

Description: (1-1119) from PCTUS9913024.seq

Perfect Score: 1119

N.A. Sequence: 1 atggaataattatgcccga.....aagaatttagaagaatttaa 1119

Comp: tacccttttaataaccggt.....ttcttaattcttccaatt

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: embl-est58

1:em\_est10 2:em\_est11 3:em\_est17 4:em\_est18 5:em\_est2  
6:em\_est9 7:em\_gss1

Database: genbank-est111

8:gb\_est1 9:gb\_est10 10:gb\_est11 11:gb\_est12 12:gb\_est13  
13:gb\_est14 14:gb\_est15 15:gb\_est16 16:gb\_est17  
17:gb\_est18 18:gb\_est19 19:gb\_est20 20:gb\_est21  
21:gb\_est22 22:gb\_est23 23:gb\_est24 24:gb\_est25  
25:gb\_est26 26:gb\_est27 27:gb\_est28 28:gb\_est29  
29:gb\_est30 30:gb\_est31 31:gb\_est32 32:gb\_est33  
33:gb\_est34 34:gb\_est35 35:gb\_est36 36:gb\_est37  
37:gb\_est38 38:gb\_est39 39:gb\_gss3 40:gb\_gss4 41:gb\_gss5 42:gb\_gss6

Statistics: Mean 11.199; Variance 2.112; scale 5.303

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.	
1	59	5.3	252	17	AA754459 97SN1787 Rice Immature	4.05e-63	
c	2	56	5.0	252	17	AA754459 97SN1787 Rice Immature	1.58e-57
3	54	4.8	247	17	AA754458 97SN1784 Rice Immature	7.72e-54	
c	4	51	4.6	247	17	AA754458 97SN1784 Rice Immature	2.26e-48
5	41	3.7	2275	20	AF034173 Human mRNA (T 7.60e-31)		
c	6	28	2.5	2275	20	AF034173 Human mRNA (T 1.70e-10)	
7	27	2.4	507	28	AI506433 vm55e09.xl Stratagene	4.19e-09	
c	8	27	2.4	595	15	AA596420 vm55e09.xl Stratagene	4.19e-09
c	9	25	2.2	369	35	AA032210 zk19d04.sl Soares.preg	2.00e-05
c	10	25	2.2	517	8	T58089 yb26a12.r1 Stratagene	2.00e-06

C	11	23	2.1	251	24	AI205893
12	23	2.1	318	37	B96187	qf60a07.xl Soares_test
13	23	2.1	325	17	AA739545	T18ESTFB TAMU Arabidop
14	23	2.1	381	41	AA739545	310 PtiFG2 Pinus taeda
15	24	2.1	383	38	AA028987	RCI11-87A4.TJ RPC111
16	24	2.1	396	10	AA234431	CJT-HSP-2313A16.TR CIT
17	23	2.1	407	35	AA056859	z172a12.r1 Soares_NHm
18	24	2.1	435	24	AI242393	SWMCA722SK Brugia mal
19	23	2.1	441	39	AQ113414	QU36A09.xl NCI_CGAP_LY
20	23	2.1	447	19	T89700	CIT-HSP-2378L7.TR CIT-
21	23	2.1	453	40	AQ227532	Y99H08.sl Soares_feta
22	23	2.1	454	41	AQ359215	HS_2019.B2.D08.MR CIT
23	23	2.1	475	39	AQ126037	HS_5029.B2.H08.SP6E RP
24	23	2.1	488	26	AI370893	HS_3036.AL.B09.MF CIT
25	23	2.1	502	9	AA151503	ta12405.xl Soares_tota
26	23	2.1	517	26	AI398575	z127B05.r1 Soares_preg
27	23	2.1	520	41	AQ334844	NCW01G1T7 Westergaards
28	23	2.1	527	17	AA651439	HS_5018.AL.F08.T7 RPCI
29	23	2.1	531	27	AI485437	31553 Lambda-PRL2 Arab
30	23	2.1	605	22	AU015321	EST243758 tomato ovary
31	24	2.1	606	41	AQ362692	AU015321 Mouse two-cel
32	23	2.1	637	33	W03126	nbxb0050006f CUGI Rice
33	24	2.1	654	39	AQ111972	z553c03.r1 Soares_feta
34	23	2.1	697	37	B16516	CIT-HSP-2377B1.TR CIT-
35	24	2.1	708	14	AA550589	342H10.TPB CIT978SKA1
36	24	2.1	764	14	AA550589	nbxb0060K14r CUGI Rice
37	24	2.1	789	41	AQ291506	1754m3 gmbPFB3.1, G.
38	23	2.1	997	39	AF010972	Hbmo0025G01f CUGI Rice
39	22	2.0	331	17	AA704953	Xbmo sapiens chromosom
40	22	2.0	360	32	D75694	z195A01.sl Soares_feta
41	22	2.0	372	15	C44854	CELK109AZF Yuji Kohara
42	22	2.0	449	23	AI175073	C44854 Yuji Kohara unp
43	22	2.0	609	25	AI294174	EST218591 Normalized r
44	22	2.0	722	37	B67669	LP07487.5Prime LP Dros
45	22	2.0	812	41	AQ291855	T23N14TR TAMU Arabidop
						nbxb0040K19f CUGI Rice

## ALIGNMENTS

RESULT 1  
LOCUS AA754459 252 bp mRNA EST 20-JAN-1998  
DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa  
ACCESSION AA754459  
NID 92801165  
VERSION AA754459.1 GI:2801165  
KEYWORDS EST.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Oryza.  
REFERENCE 1 (bases 1 to 252)  
AUTHORS Nahn,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,  
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,  
Lee,M.C. and Eun,M.Y.  
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed  
JOURNAL Unpublished (1998)  
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797457.

Contact: Eun M.Y.  
Department of Cytogenetics  
National Inst. of Agri. Sci. and Tech, RDA  
Suwon, Kyunggi-do, Korea  
Tel: 82 31 290 0301  
Fax: 82 31 290 0307  
Email: myeun@n20.osti.re.kr  
Submitted by Baek Hie Nahn, Dept of Biological Science, Myongji  
University, Yongin, Korea. 449-728 bnaahm@bioserver.myongji.ac.kr  
Seq primer: M13 Reverse Primer.  
FEATURES  
source  
1..252  
/organism="Oryza sativa"

```
/cultivar="Milyang23"  
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:  
XhoI; Directional cDNA library inserted into lambda ZAPII  
vector at 5' end with EcoRI and 3' end with Xho I site."  
/db_xref="taxon:4530"  
/map="6"  
/clone="97SN1787"  
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"  
/tissue_type="Immature Seed"  
/dev_stage="5 days after pollination"  
/lab_host="E. coli SOLR"  
BASE COUNT 5 a 21 c 12 g 35 t 179 others  
ORIGIN
```

```
Query Match 5.3%; Score 59; DB 17; Length 252;  
Best Local Similarity 11.8%; Pred. NO. 4.05e-63;  
Matches 24; Conservative 109; Mismatches 66; Indels 4; Gaps 4;  
Db 26 HGNVWCVASHGNYMNVHNCBTRGTHCDCKNVNWSMTWGTWVNVNBSGDMHYWVBVBN 84  
Qy 586 agtqtgagcaattcgtcgccagttggtggaaactaatgacgaattgcgagattgccca 645  
Db 85 TKVDVGNHTRCSRWBRVTRAHYHDYNCBBYNNNDYHMHBBMYBTGCMCTMWCWBH 144  
Qy 646 acttcagctgctataagtaattctgattataatttaattcctgagattgtct-tta 704  
Db 145 YNTKCTASGWHITSNYDVKSSTNTWGTBSYDKSMHGYWCSBBVKYHTKYSTTRATRSYT 204  
Qy 705 tcttaagctcagttatagctgctgattgattcgttctgcttggaactaca-aatatat 763  
Db 205 CVRKCYVMWMTK-KVVKYHYVVB 226  
Qy 764 ctatccgttttataatgcattct 786
```

```
RESULT 2  
LOCUS AA754459 252 bp mRNA EST 20-JAN-1998  
DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa  
CDNA clone 97SN1787, mRNA sequence.  
ACCESSION AA754459  
NID 92801165  
VERSION AA754459.1 GI:2801165  
KEYWORDS EST.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa.  
REFERENCE 1 (bases 1 to 252)  
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,  
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,  
Lee,M.C. and Eun,M.Y.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Oryza.  
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed  
JOURNAL Unpublished (1998)  
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797457.
```

```
Contact: Eun M.Y.  
Department of Cytogenetics  
National Inst. of Agri. Sci. and Tech, RDA  
Suwon, Kyunggido, Korea  
Tel: 82 331 290 0301  
Fax: 82 331 290 0307  
Email: myeun@sun20.asti.re.kr  
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji  
University, Yongin, Korea. 449-728 bhna@bioserver.myongji.ac.kr  
Seq primer: M13 Reverse Primer.  
Location/Qualifiers  
1. .252  
/organism="Oryza sativa"  
/cultivar="Milyang23"  
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:  
XhoI; Directional cDNA library inserted into lambda ZAPII  
vector at 5' end with EcoRI and 3' end with Xho I site."
```

```
FEATURES  
source  
1. .252  
/organism="Oryza sativa"  
/cultivar="Milyang23"  
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:  
XhoI; Directional cDNA library inserted into lambda ZAPII  
vector at 5' end with EcoRI and 3' end with Xho I site."
```

```
/db_xref="taxon:4530"  
/map="6"  
/clone="97SN1787"  
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"  
/tissue_type="Immature Seed"  
/dev_stage="5 days after pollination"  
/lab_host="E. coli SOLR"  
BASE COUNT 5 a 21 c 12 g 35 t 179 others  
ORIGIN
```

```
Query Match 5.0%; Score 56; DB 17; Length 252;  
Best Local Similarity 12.1%; Pred. NO. 1.58e-57;  
Matches 26; Conservative 109; Mismatches 76; Indels 3; Gaps 3;
```

```
Db 31 WVCVASHGNYMNVHNCBTRGTHCDCKNVNWSMTWGTWVNVNBSGDMHYWVBVBNTKVDVG 90  
Cp 1014 aacagcaattcattgacgcaattcagcaggtttcgaaatgcccaattctctgatac 955  
Db 91 NHTRCRSRWBRVTRAHYHDYNCBBYNNNDYHMHBB-BMYBBTGCMCTMWCWBHNTKC 149  
Cp 954 ttcagaaaagcttaattctcattcattcagcagcttgccagcttgctgctgctgata 895  
Db 150 TA-SGWHITSNYDVKSSTNTWGTBSYDKSMHGYWCSBBVKYHTKYSTTRATRSYT 208  
Cp 894 tatcgtataatcctactgcgttttctggaacactacataatctcttatttcccatga 835  
Db 209 YCVMMWTKKVVK-KYHVVBSCBHTDSCKTKTMM 241  
Cp 834 cccagcttgcagtaaatcagcttggttgaa 801
```

```
RESULT 3  
LOCUS AA754458 247 bp mRNA EST 20-JAN-1998  
DEFINITION 97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa  
CDNA clone 97SN1784, mRNA sequence.  
ACCESSION AA754458  
NID 92801164  
VERSION AA754458.1 GI:2801164  
KEYWORDS EST.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa.
```

```
REFERENCE 1 (bases 1 to 247)  
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,  
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,  
Lee,M.C. and Eun,M.Y.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Oryza.  
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed  
JOURNAL Unpublished (1998)  
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797455.
```

```
Contact: Eun M.Y.  
Department of Cytogenetics  
National Inst. of Agri. Sci. and Tech, RDA  
Suwon, Kyunggido, Korea  
Tel: 82 331 290 0301  
Fax: 82 331 290 0307  
Email: myeun@sun20.asti.re.kr  
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji  
University, Yongin, Korea. 449-728 bhna@bioserver.myongji.ac.kr  
Seq primer: M13 Reverse Primer.  
Location/Qualifiers  
1. .247  
/organism="Oryza sativa"  
/cultivar="Milyang23"  
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:  
XhoI; Directional cDNA library inserted into lambda ZAPII  
vector at 5' end with EcoRI and 3' end with Xho I site."  
/db_xref="taxon:4530"  
/map="6"  
/clone="97SN1784"
```

```
FEATURES  
source  
1. .247  
/organism="Oryza sativa"  
/cultivar="Milyang23"  
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:  
XhoI; Directional cDNA library inserted into lambda ZAPII  
vector at 5' end with EcoRI and 3' end with Xho I site."  
/db_xref="taxon:4530"  
/map="6"  
/clone="97SN1784"  
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
```



/tissue_type="Immature Seed"	
/dev_stage="5 days after pollination"	
/lab_host="E. coli SOLR"	
BASE COUNT	7 a 16 c 21 g 34 t 169 others
ORIGIN	
Query Match 4.8%; Score 54; DB 17; Length 247;	
Best Local Similarity 15.5%; Pred. No. 7.72e-54;	
Matches 35; Conservative 106; Mismatches 81; Indels 4; Gaps 3;	
Db	25 HMTBWBCCVRRYGTITNCKNGHRTTWNDCSDNAHCRTYVBWYYARSKYGYGTBY-S 83
Qy	198 ccttcacccggagcattgtagtagtaagcatcatgccttcctgactccaatgatac 257
Db	84 WNVDTNTGGTGVGKTTVNVHSGNNRCSNVVWBTAYCDYBHBXDRANWDDTRCTND 143
Qy	258 tacatcagctgcttggtagttcatgaaatctcttgacgaataatcgaattgtgctat 317
Db	144 TGCYNYTASONGTSA--TKRVGYGKTDSDCGG-GCWRKVTYSSBYBRCGVNVVVRTTS 200
Qy	318 attaactagtggaaggttaattctcctgaagtagtatcttgtagtgaacgcgcgg 377
Db	201 MWTDKSTFKMSMDMSRRSRVHYGWRMBNKKRGSRNWTDTKTWTR 246
Qy	378 aacgtctgccttcacatgctgattctatatgtcaagattgcagta 423
RESULT	4
LOCUS	AA754458 247 bp mRNA EST 20-JAN-1998
DEFINITION	97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
	cDNA clone 97SN1784, mRNA sequence.
ACCESSION	AA754458
NID	92801164
VERSION	AA754458.1 GI:2801164
KEYWORDS	EST
SOURCE	Oryza sativa.
ORGANISM	Oryza sativa
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
AUTHORS	1 (bases 1 to 247) Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.
TITLE	Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL	Unpublished (1998)
COMMENT	On Jan 14, 1998 this sequence version replaced gi:1797455.
Contact: Eun M.Y.	
Department of Cytogenetics	
National Inst. of Agri. Sci. and Tech, RDA	
Suwon, Kyunggido, Korea	
Tel: 82 331 290 0301	
Fax: 82 331 290 0307	
Email: myeun@sun20.asti.re.kr	
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhna@mbioserver.myongji.ac.kr	
Seq primer: M13 Reverse Primer.	
Location/Qualifiers	
1..247	
/organism="Oryza sativa"	
/cultivar="Milyang23"	
/note="Vector: pBluescript SK(+); Site.1: EcoRI; Site.2: XhoI; Directional cDNA library inserted into lambda ZAPII vector at 5' end with EcoRI and 3' end with Xho I site."	
/db_xref="taxon:4530"	
/map="6"	
/clone="97SN1784"	
/clone.lib="Rice Immature Seed Lambda ZAPII cDNA Library"	
/tissue_type="Immature Seed"	
/dev_stage="5 days after pollination"	
/lab_host="E. coli SOLR"	
/lab_16 c 21 g 34 t 169 others	
BASE COUNT	7 a

[illegible]



This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:56400

Seq primer: -28m13 rev1 ET from Amersham

High quality sequence stop: 446.

Location/Qualifiers

#### FEATURES

source

1. 595  
/organism="Mus musculus"  
/note="Organ: blood; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. M30 CD4+ cells. Average insert size: 1.0 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGCAGG  
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."  
/db\_xref="taxon:10090"  
/clone="IMAGE:1002184"  
/clone\_lib="Stratagene mouse Tcell 937311"  
/tissue\_type="Tcell"  
/dev\_stage="M30 CD4+ cells"  
/lab\_host="SOLR (kanamycin resistant)"  
158 a 135 c 119 g 182 t 1 others

BASE COUNT 158 a 135 c 119 g 182 t 1 others

ORIGIN

Query Match 2.4%; Score 27; DB 15; Length 595;  
Best Local Similarity 78.7%; Pred. No. 4.19e-09;  
Matches 37; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 411 CTTGTGTTGCTCAATATACATTCAGATGCTTTGTGAACACGATA 457

Cp 811 cttgtgtgaagaataatccgttagatgcattataaaacggata 765

#### RESULT

LOCUS AA032210 369 bp mRNA EST 21-AUG-1996  
DEFINITION zk19d04.s1 Soares\_pregnant\_uterus\_NbHPU Homo sapiens cDNA clone  
IMAGE:470983 3', mRNA sequence.

ACCESSION AA032210

NID G1502218

VERSION AA032210.1 GI:1502218

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 369)

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT On May 9, 1995 this sequence version replaced gi:802460.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40m13 fwd. from Amersham

High quality sequence stop: 277.

Location/Qualifiers

#### FEATURES

source

1. 369  
/organism="Homo sapiens"  
/note="Organ: uterus; Vector: pT7T3-Pac; Site\_1: Not I;  
Site\_2: Eco RI; 1st strand cDNA was primed with a Not I -  
oligo(dt) primer [5'  
AACTGGAAGATTCGCGCGCCCTTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. Library

went through one round of normalization. Library  
constructed by M. Fatima Bonaldo."  
/db\_xref="GDB:3756985"  
/db\_xref="taxon:9606"  
/map="12p"  
/clone="IMAGE:470983"  
/clone\_lib="Soares\_pregnant\_uterus\_NbHPU"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
134 a 69 c 58 g 107 t 1 others

BASE COUNT 134 a 69 c 58 g 107 t 1 others

ORIGIN

Query Match 2.2%; Score 25; DB 35; Length 369;  
Best Local Similarity 65.8%; Pred. No. 2.00e-06;  
Matches 50; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Db 246 TAAATAGCCATCAGATCAATGGAATTTTAAGATGAGATTAACATNTCCAGTATATT 305

Cp 738 taaatcagcatcagcatataactgagctttaagaataaagacatctccaggaaattaaatt 679

Db 306 ATACTCACAAAATATA 321

Cp 678 ataactcagataattta 663

#### RESULT

LOCUS T58089 517 bp mRNA EST 08-FEB-1995  
DEFINITION yb26a12.r1 Stratagene fetal spleen (#937205) Homo sapiens cDNA  
clone IMAGE:72286 5' similar to similar to gb:M31468 RAS-LIKE  
PROTEIN TC21 (HUMAN), mRNA sequence.

ACCESSION T58089

NID G659950

VERSION T58089.1 GI:659950

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 517)

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,  
Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 887

High quality sequence stops: 312 Source: IMAGE Consortium, LLNL This  
clone is available royalty-free through LLNL ; contact the IMAGE  
Consortium (info@image.llnl.gov) for further information.

Insert Length: 887 Std Error: 0.00

Seq Primer: M13Rp1

High quality sequence stop: 312.

Location/Qualifiers

source

1. 517  
/organism="Homo sapiens"  
/note="Organ: spleen; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Pooled spleens. Average insert size: 1.0 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGCAGG  
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."  
/db\_xref="GDB:49351"  
/db\_xref="taxon:9606"  
/map="17 cen-qter"  
/clone="IMAGE:72286"  
/clone\_lib="Stratagene fetal spleen (#937205)"

```
/tissue_type="fetal spleen"
/dev_stage="fetal"
/lab_host="SOLR cells (kanamycin resistant)"
BASE COUNT      130 a      98 c      135 g      145 t      9 others
ORIGIN

Query Match
Best Local Similarity 2.2%; Score 25; DB 8; Length 517;
Matches 54; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Db 98 ATGAGCTGCCATCTGGACATTTTGGATACAGCTGGAGCAGAGATTTCAGCCATGC 157
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cp 427 atgatactgcaaatcttgacatagataatcagatggaagcagacgttcgcggcttc 368

Db 158 GNGACCATGATATGAGCGCAGGAG 181
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cp 367 ttaaccaagatactactctcaggag 344

RESULT 11
LOCUS      AI205893      251 bp      mRNA      EST      16-OCT-1998
DEFINITION qf60a07.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1754388
3', mRNA sequence.
ACCESSION  AI205893
NID         93764565
VERSION     AI205893.1 GI:3764565
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 251)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On Jan 19, 1998 this sequence version replaced gi:2151858.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 245.
Location/Qualifiers
1..251
/organism="Homo sapiens"
/clone="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo.
/db_xref="taxon:9606"
/clone="IMAGE:1754388"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
BASE COUNT      95 a      34 c      27 g      95 t
ORIGIN

/tissue_type="fetal spleen"
/dev_stage="fetal"
/lab_host="SOLR cells (kanamycin resistant)"
BASE COUNT      130 a      98 c      135 g      145 t      9 others
ORIGIN

Query Match
Best Local Similarity 2.1%; Score 23; DB 24; Length 251;
Matches 41; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

Db 125 TACAGTCTCAATATGAAÁ-CTAAAAAGTTTCGATAATCATTTAAGCATTTCCT 177
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cp 543 taaactgtcaaatatacaatctaaatagtttgataatcaatctgctcttct 490

RESULT 12
LOCUS      B96187      318 bp      DNA      GSS      30-MAR-1998
DEFINITION T18E5F8B TAMU Arabidopsis thaliana genomic clone T18E5, genomic
survey sequence.
ACCESSION  B96187
NID         g2998663
VERSION     B96187.1 GI:2998663
KEYWORDS    GSS.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 318)
AUTHORS   Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and
Venter,J.C.
TITLE     A BAC End Sequence Database for Identifying Minimal Overlaps in
Arabidopsis Genomic Sequencing. Update 3
JOURNAL   Unpublished (1997)
COMMENT   Other_GSS: T18E5TR
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M13-21
Class: BAC ends
High quality sequence stop: 318.
Location/Qualifiers
1..318
/organism="Arabidopsis thaliana"
/strain="Columbia"
/clone="vector: BelobACII; Site_1: HindIII; Site_2:
HindIII; Produced by Rod Wing"
/d_xref="taxon:3702"
/clone="T18E5"
/clone_lib="TAMU"
/sex="hermaphrodite"
BASE COUNT      93 a      42 c      51 g      131 t      1 others
ORIGIN

Query Match
Best Local Similarity 2.1%; Score 23; DB 37; Length 318;
Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 204 ATAATCGAATTTTCTGCTATAACTA 230
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 299 ataactcgatgtgtgctataacta 325

RESULT 13
LOCUS      AA739545      325 bp      mRNA      EST      15-JAN-1998
DEFINITION 310 PfIFG2 Pinus taeda cDNA clone 8507M 3', mRNA sequence.
ACCESSION  AA739545
NID         g2778096
VERSION     AA739545.1 GI:2778096
KEYWORDS    EST.
SOURCE      loblolly pine.
ORGANISM    Pinus taeda
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;
```

[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)

```

Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..381
/organism="Homo sapiens"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC Library"
/db_xref="GDB:7533027"
/db_xref="taxon:9606"
/clone="R-87A4"
/clone_lib="RPC11"
/sex="Male"
/cell_type="Lymphocytes"
BASE COUNT      121 a   56 c   64 g   140 t
ORIGIN
Query Match      2.1%; Score 23; DB 41; Length 381;
Best Local Similarity 79.5%; Pred. No. 6.63e-04;
Matches 31; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 286 TATGGTAAAGATTCTTATTCATCATGTGCTGTTTATA 324
||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 404 tattgtcaagattgcgcatcattatgctgtttata 442

RESULT 15
LOCUS          AQ028987       383 bp      DNA           GSS             29-JUN-1998
DEFINITION     CIT-HSP-2313A16.TR CIT-HSP Homo sapiens genomic clone 2313A16,
               genomic survey sequence.
ACCESSION      AQ028987
NID            93269209
VERSION        AQ028987.1 GI:3269209
KEYWORDS       GSS.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
               Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 383)
AUTHORS        Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
               Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
               Simon,M., and Venter,J.C.
TITLE          Use of a random BAC End Sequence Database for Sequence-Ready Map
               Building
JOURNAL        Unpublished (1997)
COMMENT        Other_GSSs: CIT-HSP-2313A16.TF
               Contact: Mark Adams
               Department of Eukaryotic Genomics
               The Institute for Genomic Research
               9712 Medical Center Dr., Rockville, MD 20850, USA
               Tel: 301 838 0200
               Fax: 301 838 0208
               Email: mdadams@tigr.org
               Clones are available from Research Genetics (info@resgen.com). BAC
               end search page:
               http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
               Seq primer: M13 Reverse
               Class: BAC ends.
FEATURES       Location/Qualifiers
source         1..383
               /organism="Homo sapiens"
               /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
               HindIII"
               /db_xref="taxon:9606"
               /clone="2313A16"
               /clone_lib="CIT-HSP"
               /sex="Male"
               /cell_type="Sperm"
BASE COUNT      127 a   74 c   50 g   132 t
ORIGIN
Query Match      2.1%; Score 24; DB 38; Length 383;
Best Local Similarity 71.4%; Pred. No. 3.82e-05;

```

---

```

Pinaceae; Pinus.
1 (bases 1 to 325)
Kinlaw,C.S.
LOBLOLLY pine CDNAS
Unpublished (1995)
On Jan 7, 1998 this sequence version replaced gi:948055.

Contact: Claire S. Kinlaw
USDA IFG Dendrome Project
Institute of Forest Genetics
Dendrome Project, Institute of Forest Genetics, P.O. Box 245,
Berkeley, CA 94701
Tel: 5105596429
Fax: 5105596440
Email: cskes2w007.pswfs.gov
The sequence entry for this EST has been reverse complimented and
is being submitted in the sense orientation.
Seq primer: M13 Universal.
Location/Qualifiers
1..325
/organism="Pinus taeda"
/note="Vector: lambda ZAP; Site_1: EcoRI; Site_2: XhoI;
The tissue source for this library is xylem. The xylem
tissue was harvested in spring and summer from branches
of seed orchard trees which are clones of the same
genotype. Branches were 4-6 inches in diameter. The cdnas
were directionally cloned into Lambda Zap and were
rescued as a Bluescript derivative in the EcoRI and XhoI
sites."
/db_xref="taxon:3352"
/clone="R507M"
/clone_lib="PtiFG2"
/tissue_type="xylem"
BASE COUNT      93 a   41 c   63 g   128 t
ORIGIN
Query Match      2.1%; Score 23; DB 17; Length 325;
Best Local Similarity 69.5%; Pred. No. 6.63e-04;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Db 249 TTTCGCTGTAATCTCGGAATGGAAGTTGAAGCAATTAATATGCATGTTTTCCGAAAA 307
||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 817 tttaetggcaactggctcggaatgaatgaagaagtattgttggtccagaaaa 875

RESULT 14
LOCUS          AQ284757       381 bp      DNA           GSS             22-NOV-1998
DEFINITION     RPC111-87A4.TJ RPC111 Homo sapiens genomic clone R-87A4, genomic
               survey sequence.
ACCESSION      AQ284757
NID            93910997
VERSION        AQ284757.1 GI:3910997
KEYWORDS       GSS.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
               Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 381)
AUTHORS        Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
               Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P., and Venter,J.C.
TITLE          Use of human BAC End Sequences for Sequence-Ready Map Building
               Unpublished (1998)
COMMENT        Contact: Mark Adams
               Department of Eukaryotic Genomics
               The Institute for Genomic Research
               9712 Medical Center Dr., Rockville, MD 20850, USA
               Tel: 301 838 0200
               Fax: 301 838 0208
               Email: mdadams@tigr.org
               For clone availability, please contact Pieter de Jong
               (pieter@dejong.med.buffalo.edu). Clones may be purchased from
               Research Genetics (info@resgen.com). BAC end search page:

```

Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 326 CTAATTCAGAAGCAAAATTCACATTTTACTATCCACTTTAGTTACCATTCCTTA 381

Cp 355 ctacttcaggaggaaaatttaacccttccactagtttaataatagcaacaattcgatta 300

Search completed: Thu Sep 2 14:51:10 1999

Job time : 2695 secs.

\*\*\*\*\*  
 M E R E L L  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
 Copyright (C) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 14:51:32 1999; MasPar time 249.17 Seconds  
 Tabular output not generated. 962.282 Million cell updates/sec

Title: >PCT-US99-13024-1  
 Description: (1-1119) from PCTUS9913024.seq  
 Perfect Score: 1119  
 N.A. Sequence: 1 atggaataatttatggccga.....aagaatttagagaagtttaa 1119  
 Comp: taccttttaataccggct.....ttcttaatactctcaatt

Scoring table: TABLE default  
 Gap 6

Nmatch STD : Dbase 0: Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: n-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21 22:part22 23:part23  
 24:part24 25:part25 26:part26 27:part27 28:part28  
 29:part29 30:part30 31:part31 32:part32 33:part33  
 34:part34 35:part35 36:part36 37:part37 38:part38  
 39:part39 40:part40 41:part41 42:part42 43:part43  
 44:part44 45:part45 46:part46 47:part47 48:part48  
 49:part49 50:part50 51:part51 52:part52 53:part53  
 54:part54 55:part55 56:part56 57:part57 58:part58  
 59:part59 60:part60

Statistics: Mean 8.996; Variance 5.350; scale 1.682

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.	
1	1078	96.3	8855	24	Phage T4 tail fibre g	0.00e+00	
2	39	3.5	91	Q51746	Oligonucleotide probe	8.60e-09	
3	38	3.4	204	N81164	Base substituted E.co	3.23e-08	
4	37	3.3	91	Q51746	Oligonucleotide probe	1.20e-07	
5	37	3.3	204	N81164	Base substituted E.co	4.41e-07	
6	36	3.2	91	V44650	Mammalian DNA replica	2.50e-04	
7	31	2.8	114	12	Q70465	Generic DNA sequence	8.57e-04
8	30	2.7	91	V44650	Mammalian DNA replica	8.57e-04	
9	30	2.7	114	12	Q70468	Generic DNA sequence	8.57e-04

10	29	2.6	114	12	Q70469	Generic DNA sequence	2.89e-03
11	29	2.6	114	12	Q70466 <td>Generic DNA sequence <td>2.89e-03</td> </td>	Generic DNA sequence <td>2.89e-03</td>	2.89e-03
12	29	2.6	114	12	Q70467 <td>Generic DNA sequence <td>2.89e-03</td> </td>	Generic DNA sequence <td>2.89e-03</td>	2.89e-03
13	29	2.6	114	12	Q70469 <td>Generic DNA sequence <td>2.89e-03</td> </td>	Generic DNA sequence <td>2.89e-03</td>	2.89e-03
14	28	2.5	114	12	Q70466 <td>Generic DNA sequence <td>9.59e-03</td> </td>	Generic DNA sequence <td>9.59e-03</td>	9.59e-03
15	27	2.4	39	7	Q51787 <td>Mixed oligonucleotide <td>3.13e-02</td> </td>	Mixed oligonucleotide <td>3.13e-02</td>	3.13e-02
16	27	2.4	114	12	Q70467 <td>Generic DNA sequence <td>3.13e-02</td> </td>	Generic DNA sequence <td>3.13e-02</td>	3.13e-02
17	27	2.4	114	12	Q70465 <td>Generic DNA sequence <td>3.13e-02</td> </td>	Generic DNA sequence <td>3.13e-02</td>	3.13e-02
18	27	2.4	114	12	Q70465 <td>Generic DNA sequence <td>3.13e-02</td> </td>	Generic DNA sequence <td>3.13e-02</td>	3.13e-02
19	27	2.4	501	3	N50026 <td>Sequence encoding new</td> <td>3.13e-02</td>	Sequence encoding new	3.13e-02
20	26	2.3	39	7	Q51787 <td>Mixed oligonucleotide</td> <td>1.00e-01</td>	Mixed oligonucleotide	1.00e-01
21	26	2.3	114	12	Q70468 <td>Generic DNA sequence <td>1.00e-01</td> </td>	Generic DNA sequence <td>1.00e-01</td>	1.00e-01
22	25	2.2	498	3	N50034 <td>Sequence encoding new</td> <td>3.15e-01</td>	Sequence encoding new	3.15e-01
23	25	2.2	501	3	N50030 <td>Sequence encoding new</td> <td>3.15e-01</td>	Sequence encoding new	3.15e-01
24	25	2.2	501	3	N50023 <td>Sequence encoding new</td> <td>3.15e-01</td>	Sequence encoding new	3.15e-01
25	25	2.2	501	3	N50032 <td>Sequence encoding new</td> <td>3.15e-01</td>	Sequence encoding new	3.15e-01
26	25	2.2	501	3	N50029 <td>Sequence encoding new</td> <td>3.15e-01</td>	Sequence encoding new	3.15e-01
27	25	2.2	501	3	N50027 <td>Sequence encoding new</td> <td>3.15e-01</td>	Sequence encoding new	3.15e-01
28	25	2.2	501	3	N50031 <td>Sequence encoding new</td> <td>3.15e-01</td>	Sequence encoding new	3.15e-01
29	25	2.2	501	3	N50025 <td>Sequence encoding new</td> <td>3.15e-01</td>	Sequence encoding new	3.15e-01
30	25	2.2	501	3	N50033 <td>Sequence encoding new</td> <td>3.15e-01</td>	Sequence encoding new	3.15e-01
31	25	2.2	501	3	N50028 <td>Sequence encoding new</td> <td>3.15e-01</td>	Sequence encoding new	3.15e-01
32	25	2.2	501	3	N50025 <td>Sequence encoding new</td> <td>3.15e-01</td>	Sequence encoding new	3.15e-01
33	24	2.1	68	34	T73397 <td>Oligonucleotide tag c</td> <td>9.69e-01</td>	Oligonucleotide tag c	9.69e-01
34	24	2.1	68	32	T63255 <td>Messenger RNA primer</td> <td>9.69e-01</td>	Messenger RNA primer	9.69e-01
35	24	2.1	114	12	Q70472 <td>Generic DNA sequence</td> <td>9.69e-01</td>	Generic DNA sequence	9.69e-01
36	24	2.1	114	12	Q70470 <td>Generic DNA sequence</td> <td>9.69e-01</td>	Generic DNA sequence	9.69e-01
37	24	2.1	498	3	N50034 <td>Sequence encoding new</td> <td>9.69e-01</td>	Sequence encoding new	9.69e-01
38	24	2.1	501	3	N50028 <td>Sequence encoding new</td> <td>9.69e-01</td>	Sequence encoding new	9.69e-01
39	24	2.1	501	3	N50029 <td>Sequence encoding new</td> <td>9.69e-01</td>	Sequence encoding new	9.69e-01
40	24	2.1	501	3	N50024 <td>Sequence encoding new</td> <td>9.69e-01</td>	Sequence encoding new	9.69e-01
41	24	2.1	501	3	N50031 <td>Sequence encoding new</td> <td>9.69e-01</td>	Sequence encoding new	9.69e-01
42	24	2.1	501	3	N50023 <td>Sequence encoding new</td> <td>9.69e-01</td>	Sequence encoding new	9.69e-01
43	24	2.1	501	3	N50026 <td>Sequence encoding new</td> <td>9.69e-01</td>	Sequence encoding new	9.69e-01
44	24	2.1	501	3	N50032 <td>Sequence encoding new</td> <td>9.69e-01</td>	Sequence encoding new	9.69e-01
45	24	2.1	855	60	X13639 <td>Enterococcus faecalis</td> <td>9.69e-01</td>	Enterococcus faecalis	9.69e-01

## ALIGNMENTS

RESULT 1  
 ID T29053 standard; DNA; 8855 BP.  
 AC T29053;  
 DE 07-JAN-1997 (first entry)  
 DT Phage T4 tail fibre genes.  
 KW Phage T4; tail fibre protein; nanotechnology; nano-structure;  
 OS filter; molecular sieve; ss.  
 OS Bacteriophage T4.  
 FH Key Location/Qualifiers  
 FT cds 16..3885  
 FT /\*tag= a  
 FT /label= Gene\_34  
 FT /product= gp34 protein  
 FT 3894..4091  
 FT /\*tag= b  
 FT /label= ORFX  
 FT /product= unidentified product  
 FT 4127..5014  
 FT /\*tag= c  
 FT /label= Gene\_35  
 FT /product= gp35 protein  
 FT 5077..5742  
 FT /\*tag= d  
 FT /label= Gene\_36  
 FT /product= gp36 protein  
 FT 5751..8831  
 FT /\*tag= e  
 FT /label= Gene\_37  
 FT /product= gp37 protein  
 PN WP9611947-A1.  
 PD 25-APR-1996.  
 PF 13-OCT-1995; UI3023.  
 PR 13-OCT-1994; US-322760.





```

OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT /tag= a
FT /function= multiple cloning site
FT primer_bind 187..204
FT /tag= b
PD EP-285123-A.
PN 05-MAY-1988.
PR 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivu A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prep of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; P; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 3.4%; Score 38; DB 1; Length 204;
Best Local Similarity 11.2%; Pred. No. 3.23e-08;
Matches 11; Conservative 50; Mismatches 37; Indels 0; Gaps 0;

Db 88 rtthhyrmbvdyrdnsdaawccyrrsvkdyccynachdhdybbvbyvnhn 147
Cp :||| : : : : : : : : : : : : : : : : : : : : : : : : :
759 attgttagttccaggaagtaataatcagcatataactgagcttaagataaag 700
Db 148 hnnccnccbnhchvbnhnrwnayvrhddvhh 185
Cp :||| : : : : : : : : : : : : : : : : : : : : : : : : :
699 aacatctccagggaataataatcagataattac 662

RESULT 4
ID Q51746 standard; cDNA; 91 BP.
AC Q51746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT ) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR WPI: 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 3.3%; Score 37; DB 9; Length 91;
Best Local Similarity 10.5%; Pred. No. 1.20e-07;
Matches 11; Conservative 57; Mismatches 45; Indels 0; Gaps 0;

Db 73 aayvcchvccgymttthhyrmbvdyrdnsdaawccyrrsvkdyccynachh 132
QY :||| : : : : : : : : : : : : : : : : : : : : : : : : :
226 aagtcatacgcttttcgacttcaaatgatactacatcagctgtttgttagtcatg 285
Db 133 ddhyvbybbvynvnhnncnccbnhchvbnhnrwnayvrhddvhh 185
QY :||| : : : : : : : : : : : : : : : : : : : : : : : : :
286 aattcttcagcaataatgcaattgtgtctattataatcagtaggaagggttaa 338

RESULT 6
ID V44650 standard; DNA; 91 BP.
AC V44650;
DT 06-OCT-1998 (first entry)
DE Mammalian DNA replication origin consensus sequence, uniorsconsensus.
KW DNA replication origin; human; mammal; alphaconsensus; uniorsconsensus;
KW anti-gene; DNA replication inhibitor; shuttle vector construct creation;
KW gene therapy; ss.
OS Mammalia.
PN WO9827200-A2.
PD 25-JUN-1998.
PF 12-DEC-1997; CA0972.
PR 21-MAY-1997; US-047322.
PR 16-DEC-1996; US-033374.
PA (UYMC-) UNIV MCGILL.

```

```

Matches 6; Conservative 41; Mismatches 10; Indels 0; Gaps 0;

Db 2 gctccgcgsgsvshvyyvvhvshhhvhhvshvvhvvhvvhvvhvvhvvhv 58
FT :||| : : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 gcgaggccctcccttaagagttttgacctccaccgagcattagttagta 226

RESULT 5
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT /tag= a
FT /function= multiple cloning site
FT primer_bind 187..204
FT /tag= b
PD EP-285123-A.
PN 05-MAY-1988.
PR 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivu A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prep of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; P; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 3.3%; Score 37; DB 1; Length 204;
Best Local Similarity 9.7%; Pred. No. 1.20e-07;
Matches 11; Conservative 57; Mismatches 45; Indels 0; Gaps 0;

Db 73 aayvcchvccgymttthhyrmbvdyrdnsdaawccyrrsvkdyccynachh 132
QY :||| : : : : : : : : : : : : : : : : : : : : : : : : :
226 aagtcatacgcttttcgacttcaaatgatactacatcagctgtttgttagtcatg 285
Db 133 ddhyvbybbvynvnhnncnccbnhchvbnhnrwnayvrhddvhh 185
QY :||| : : : : : : : : : : : : : : : : : : : : : : : : :
286 aattcttcagcaataatgcaattgtgtctattataatcagtaggaagggttaa 338

RESULT 6
ID V44650 standard; DNA; 91 BP.
AC V44650;
DT 06-OCT-1998 (first entry)
DE Mammalian DNA replication origin consensus sequence, uniorsconsensus.
KW DNA replication origin; human; mammal; alphaconsensus; uniorsconsensus;
KW anti-gene; DNA replication inhibitor; shuttle vector construct creation;
KW gene therapy; ss.
OS Mammalia.
PN WO9827200-A2.
PD 25-JUN-1998.
PF 12-DEC-1997; CA0972.
PR 21-MAY-1997; US-047322.
PR 16-DEC-1996; US-033374.
PA (UYMC-) UNIV MCGILL.

```



```
Query Match      2.7%; Score 30; DB 46; Length 91;
Best Local Similarity 17.1%; Pred. No. 8.57e-04;
Matches 12; Conservative 38; Mismatches 20; Indels 0; Gaps 0;

Db 21 wgakwkwkwwhrassacmdwkaaktwkggtwaryykgkrkmwkwkawsdatakwk 80
Cp 807 tttggaagaataatccgtagatgattataaaacggatagatatgtgtgctcc 748
Db 81 dakwkmwrkt 90
Cp 747 aagagcaagt 738

RESULT 9
ID Q70468 standard; DNA: 114 BP.
AC Q70468;
DT 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /tag= a
FT /note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN W09418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UUNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI; 94-279739/34.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compsns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need
CC for complex methods of hybridoma formation or in vivo antibody
CC production. The TSARs are easily characterised and have designed activity
CC allowing direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match      2.7%; Score 30; DB 12; Length 114;
Best Local Similarity 3.6%; Pred. No. 8.57e-04;
Matches 4; Conservative 31; Mismatches 77; Indels 0; Gaps 0;

Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnn 62
Cp 305 gaattgtgctataactagtgaaagggttaatttctctcctgaagtattgttgg 364
```

```
Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnn 114
Qy 365 taagaacgcgcggacgtcgtccttccatctgattctattgttcaagatt 416

RESULT 10
ID Q70469 standard; DNA: 114 BP.
AC Q70469;
DT 07-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /tag= a
FT /note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN W09418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UUNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI; 94-279739/34.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC Q70469 is a generic DNA sequence used to generate random TSAR peptide
CC This generic formula can be represented as follows: X(TGC)(NNB)10-
CC (TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking restriction
CC sites (X is not the same as Y) that are not specified further. This
CC sequence generates peptides that are cloverleaf in structure. Other
CC generic sequences are shown in Q70465-68. Other specific peptides
CC generated by these generic sequences are shown in R65150-54. TSARs are
CC concatenated heterofunctional proteins or peptides, comprising at least
CC two functional regions - a binding domain with affinity for a ligand and
CC a second effector peptide portion that is chemically or biologically
CC active. They may further comprise a linker peptide between the 2 domains.
CC The oligonucleotides are also designed so that the expressed peptide
CC contains 2 or 4 cysteine residues positioned in, or flanking, the
CC conformational rigidity to the peptides. These residues confer some degree of
CC informational rigidity to the peptides. The TSARs or compsns. comprising
CC a TSAR binding domain can be used in vivo to deliver a chemically or
CC biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
CC or enzyme, to the specific target or on the cell. They can also replace
CC the function of macromolecules, eg. monoclonal or polyclonal antibodies
CC and therefore circumvent the need for complex methods of hybridoma
CC formation or in vivo antibody production. The TSARs are easily
CC characterised and have designed activity allowing direct and rapid
CC detection in a screening process.
SQ Sequence 114 BP; 0 A; 4 G; 4 C; 4 T;

Query Match      2.6%; Score 29; DB 12; Length 114;
Best Local Similarity 6.4%; Pred. No. 2.89e-03;
Matches 7; Conservative 28; Mismatches 74; Indels 0; Gaps 0;

Db 6 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnn 65
Qy 356 tatcttggttaagaacgcgcggacgtcgtccttccatctgattctattgttcaagatt 415
Db 66 btgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 114
Qy 416 ttgacgtatcatgtctgtctttttattacttcttcttctaaagagctatcgc 464

RESULT 11
```







\*\*\*\*\*

W O R L D  
(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_n n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Thu Sep 2 12:58:29 1999; MasPar time 2053.96 Seconds  
Tabular output not generated. 1510.140 Million cell updates/sec

Title: >PCT-US99-13024-1  
Description: (1-1119) from PCTUS9913024.seq  
Perfect Score: 1119  
N.A. Sequence: 1 atggaaataattatgcccga.....aagaatttagaagatttaa 1119  
Comp: taccttttaataaccgct.....ttcttaatactcttcaaat

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb158

1:em\_bal 2:em\_ba2 3:em\_fun 4:em\_htg 5:em\_hum1 6:em\_hum2  
7:em\_in 8:em\_om 9:em\_of 10:em\_ov 11:em\_pat 12:em\_ph  
13:em\_pl 14:em\_ro 15:em\_sts 16:em\_v1

Database: genbank11  
17:gb\_bal 18:gb\_ba2 19:gb\_htg1 20:gb\_htg2 21:gb\_in1  
22:gb\_in2 23:gb\_om 24:gb\_ov 25:gb\_pat 26:gb\_ph 27:gb\_pl1  
28:gb\_pl2 29:gb\_pr1 30:gb\_pr2 31:gb\_pr3 32:gb\_ro  
33:gb\_st 34:gb\_sts 35:gb\_sy 36:gb\_un 37:gb\_v1

Statistics: Mean 10.969; Variance 5.248; scale 2.090

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	176	15.7	4418 26	PT4G36G37 Bacteriophage T4 genes	4.71e-119
2	176	15.7	4418 26	Myr4TL Phage T4 tail fiber ge	4.71e-119
3	156	13.9	902 26	MYOX2F36 Bacteriophage OX2 tail	4.83e-102
4	152	13.6	911 26	MYR2F36 Bacteriophage T2 tail	1.17e-98
5	152	13.6	944 26	MYK3F36 Bacteriophage K3 tail	1.17e-98
6	50	4.5	7218 25	I66494 Sequence 14 from paten	1.40e-16
7	44	3.9	7218 25	I66494 Sequence 14 from paten	2.46e-12
8	37	3.3	10772 21	AF012089 Drosophila melanogaste	1.32e-07
9	37	3.3	10772 21	AF012089 Drosophila melanogaste	1.32e-07
10	34	3.0	354 21	OFU89259 Oxytricha fallax 57kd	1.13e-05
11	32	2.9	215 25	I28278 Sequence 5 from paten	2.01e-04
12	33	2.9	370 21	OFU89262 Oxytricha fallax 57kd	4.81e-05
13	33	2.9	965 25	AR024229 Sequence 22 from paten	4.81e-05

C	14	33	2.9	216021	31	HUAC004787	Homo sapiens Chromosom	4.81e-05
	15	31	2.8	216021	31	HUAC004787	Homo sapiens Chromosom	8.22e-04
C	16	30	2.7	74371	31	AC005369	Homo sapiens Chromosom	3.29e-03
C	17	29	2.6	1056	23	MYU87256	Mustela vison GT dinuc	1.29e-02
C	18	29	2.6	74371	31	AC005369	Homo sapiens chromosom	1.29e-02
C	19	26	2.3	354	21	OFU89259	Oxytricha fallax 57kD	6.61e-01
C	20	26	2.3	641	23	MYU87253	Mustela vison GT dinuc	6.61e-01
C	21	26	2.3	965	25	AR024229	Sequence 22 from paten	6.61e-01
C	22	26	2.3	7207	22	CELT35E4	Caenorhabditis elegans	6.61e-01
C	23	26	2.3	79073	27	AB008271	Arabidopsis thaliana g	6.61e-01
C	24	26	2.3	140306	19	HS1068F16	Human DNA sequence **	6.61e-01
C	25	26	2.3	158166	19	AC005813	Drosophila melanogaste	6.61e-01
C	26	25	2.2	23636	21	CELF47C10	Caenorhabditis elegans	2.32e+00
C	27	25	2.2	86004	29	HS1J723	Human DNA sequence fro	2.32e+00
C	28	24	2.1	215	25	I28278	Sequence 5 from patent	7.88e+00
C	29	24	2.1	886	31	AF084462	Homo sapiens GTP-bind1	7.88e+00
C	30	24	2.1	942	30	HSU71203	Human rit mRNA, comple	7.88e+00
C	31	24	2.1	1037	30	HSU78165	Human small G protein	7.88e+00
	32	24	2.1	1056	23	MYU87256	Mustela vison GT dinuc	7.88e+00
C	33	24	2.1	1112	30	HSRITGENE	H.sapiens mRNA for RIT	7.88e+00
	34	24	2.1	1189	17	CLOORFCD	C.difficile DNA fragme	7.88e+00
	35	24	2.1	5105	17	CDTOXUTX	C.difficile toxB and t	7.88e+00
	36	24	2.1	9770	17	CDFOXA	Clostridium difficile	7.88e+00
	37	24	2.1	10068	32	D89593	unannotatable data.	7.88e+00
C	38	24	2.1	26039	17	CD26039	C.difficile cdu2, cdu1	7.88e+00
C	39	24	2.1	31812	27	SPUNK4	S.pombe chromosome I c	7.88e+00
	40	24	2.1	37685	30	HSAC000371	Human cosmid g1980a181	7.88e+00
	41	24	2.1	43309	30	AC002498	Human Cosmid g5129g124	7.88e+00
	42	24	2.1	43407	31	AC005584	Homo sapiens clone UWG	7.88e+00
C	43	24	2.1	115905	29	HS57E3	Homo sapiens DNA seque	7.88e+00
C	44	24	2.1	137557	31	AC005909	Homo sapiens Chromosom	7.88e+00
C	45	23	2.1	156378	20	AC006453	Homo sapiens clone DJ0	2.58e+01

#### ALIGNMENTS

RESULT	1	PT4G36G37	4418 bp	DNA	PHG	15-SEP-1990
LOCUS		Bacteriophage T4 genes 36 and 37 (complete); 35 and 38 (partial)				
DEFINITION		Encoding tail fiber proteins.				
ACCESSION		J02509 J02508				
NID		G215884				
VERSION		J02509.1 GI:215884				
KEYWORDS		tail fiber protein.				
SOURCE		Bacteriophage T4 DNA, clones SA-1,SA-2,SB,SC,SD,SE-1,SE-2 and SF.				
ORGANISM		Bacteriophage T4				
REFERENCE		1 (bases 1 to 4418)				
AUTHORS		Oliver,D.B. and Crowther,R.A.				
TITLE		DNA sequence of the tail fibre genes 36 and 37 of bacteriophage T4				
JOURNAL		J. Mol. Biol. 153, 545-568 (1981)				
MEDLINE		82170495				
REFERENCE		2 (bases 180 to 235)				
AUTHORS		Christensen,A.C. and Young,E.T.				
TITLE		T4 late transcripts are initiated near a conserved DNA sequence				
JOURNAL		Nature 299, 369-371 (1982)				
MEDLINE		82272427				
COMMENT		Tail fiber genes gp36 and gp37 code for the distal half of the tail fiber of phage T4. The thin tip of the distal half fiber interacts with the bacterial lipopolysaccharide receptor and specifies the host range of the phage. The r at position 4235 in [1] should be a g in order to encode glutamic acid at that codon as stated in [1]. Location/Qualifiers 1. .4418 /organism="Bacteriophage T4" /db_xref="taxon:10665" /note="tail fiber protein 35" /codon_start=2 /transl_table=11 /protein_id="AAA32512.1"				
FEATURES		source				
		CDS				

```
/db_xref="PID:g215885"
/db_xref="GI:215885"
/translation="SFSEVNRNGGISKPAEFVNGIRVNYICESASPPDIMVLPTQAS
STKGKVGQGFREV"
241..906
/notes="tail fiber protein 36"
/codon_start=1
/transl_table=11
/protein_id="AAA32513.1"
/db_xref="PID:g215886"
/db_xref="GI:215886"
/translation="MADLKVGSTTGGSVIWHOGNFPNPGADDDVLKSKYIYSEYNKP
QAADNDFVSKANGTYASKVTNAGIYAPNIPMPCGIYGGNGDGAIFDNRKIDIV
SWYGVGKSFSGSTGRTVINTRNGDINTKGVVSAAGQVRSGAAAIANDLTKRDIY
DGAINTVTANANRSLRSGDTMTGNLTAPNFFSQNPASQPSHVPRFDQIVKDSVQDF
GYV"
915..3995
/notes="tail fiber protein 37"
/codon_start=1
/transl_table=11
/protein_id="AAA32514.1"
/db_xref="PID:g215887"
/db_xref="GI:215887"
/translation="MATLKOIQFKRSKIAGTRPAASVLAEGELAINLKDRTIFTKDDSD
GNIIDGFAKGGQVQDGNVTINGLLRLNGDYVOTGGTVNGPIGSTDGVTKIFRSTQG
SFYARATNTSNAHLWFENADGTERGVIIYARPTTDDGEIRLVRROGTGTANSEFYF
RSINGGEFOANRLASDSLVTKRIADVTVIHDAKAFQYDHSVLNYYVPGTNGV
NVLKRVAKSGGTIYHEIVTAQTLADEVSWNSGDPVFKLYGRDDGRMIIRNSLAL
GFTTFNPSDGVNGVMGDKYLVLGDTVGLSYKATGTVDFLGVGGYSVASITPDSFR
STRKGI FGSEDDGATWMPGTNAALLSVOTQADNNAGDQTHIGYNAGGKNNHYFR
GTGMINITQOQMEINPGILKLVTSNNVOYFADGTISSIQPKLNDIEFLTKSNNTA
GLKFAPAGKPTDRTIQWNGGTREGQKNVYIIKAWGNSFNATGDRSRETVFQVSDSQ
GYFYAHRKAPTGTIGRIEAFQAGDVYKAGIIANGNFRVYVSSALAGNVTMSNGLF
VOGGSITGVQIGGTANALRWNAEYGAIFRSESNFYIIPNONGESGDIHSSLR
PVRIGLNDGMVGLGRDSFVQDNNALTTINSNRINANFRMLQGSAYIDAECTDAVR
PAGASFSQNNEDVRAPYMNIDRTDASAYVILKQRYVQNGGYSGLTLLNNGNFR
VYHGGGNGSTGPOTADGWEFIKNGDFISPRDLIAGRVFRDRTGNITGGSGNFANL
NSTIESLKTDMSSYPGAPIPWPSDVPAGFALMEGOTFDKSAYPKLAVAYPSGVIP
DMRGOTIKGKPSGRAVLSAEADGVKAHSHSASASTDLGKTTSSTFDYCTKGTNSTGG
HTHSGSGSTNGEHSYIEAWNGTGVGNGKMSVYIYRAGGNTNAGNHSHTFSE
GTSSAGDHSVSGIGAHHTVAIGSHGHTIITVNSGTENTVKNIAFNIVRLA"
4022..>4418
/notes="tail fiber protein 38"
/codon_start=1
/transl_table=11
/protein_id="AAA32515.1"
/db_xref="PID:g553025"
/db_xref="GI:553025"
/translation="MKIYHYFDTKFYKEENYKPKVKGILGLPAHSTIKKPLEPKEGYA
VVFDETDQIYEEEDHRGRAMTFNKEEIFISDIGSPVGITFDEPGEFDIWDGWEKE
DEYKRLVLRNRIELEYKEFOVLNNMIEA"
BASE COUNT 1347 a 799 c 976 g 1296 t
ORIGIN 1 bp upstream of HindIII site; 153.4 kb on genomic map.

Query Match 15.7%; Score 176; DB 26; Length 4418;
Best Local Similarity 99.4%; Pred. No. 4.71e-119;
Matches 177; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 AACGTTTCTCGAGTATCAGAAATCGCGCATTCGAAACCTGCTGAATTTGGCGTCAA 60
|||||
QY 942 aagctttctcgaagtacgaagaatggcgcatttcgaaacctgctgaattggcgtcaa 1001
|||||

Db 61 TGGTATTCTGTTTAATATATATCTCGCAATCCGCTTCACCTCCGGATATATGCTACTTCC 120
|||||
QY 1002 tggattcgtgtaattatctcgaaatcgcttcacccccggatataatgctacttc 1061
|||||

Db 121 TACGCAAGCATCTCTAAACTGGTAAAGTGTGGCAGAAATTTAGAGAACTTTAA 178
|||||
QY 1062 tacgcaagcatcgtctaaactgtaagtggttggtggcagaatttagagaaagttaa 1119
|||||

RESULT 2 MYT4TL 4418 bp DNA PHG 07-JUL-1995
```

```
DEFINITION Phage T4 tail fiber genes (numbers 36, 37 and 38).
ACCESSION V00863 J02508 J02509
NID g15371
VERSION V00863.1 GI:15371
KEYWORDS coliphage T4.
SOURCE coliphage T4.
ORGANISM Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
T4-like phages.
REFERENCE 1 (bases 1 to 4418)
AUTHORS Oliver,D.B. and Crowther,R.A.
TITLE DNA sequence of the tail fibre genes 36 and 37 of bacteriophage T4
JOURNAL J. Mol. Biol. 153 (3), 545-568 (1981)
MEDLINE 82170495
FEATURES Location/Qualifiers
source 1..4418
organism="coliphage T4"
db_xref="taxon:10665"
<1..178
/notes="reading frame ? (1 is 3rd base in codon)"
/codon_start=2
/transl_table=11
/protein_id="CAA24226.1"
/db_xref="PID:g836604"
/db_xref="GI:836604"
/db_xref="SWISS-PROT:P03742"
/translation="SFSEVNRNGGISKPAEFVNGIRVNYICESASPPDIMVLPTQAS
STKGKVGQGFREV"
241..906
/notes="gene 36"
/codon_start=1
/transl_table=11
/protein_id="CAA24227.1"
/db_xref="PID:g15373"
/db_xref="GI:15373"
/db_xref="SWISS-PROT:P03743"
/translation="MADLKVGSTTGGSVIWHOGNFPNPGADDDVLKSKYIYSEYNKP
QAADNDFVSKANGTYASKVTNAGIYAPNIPMPCGIYGGNGDGAIFDNRKIDIV
SWYGVGKSFSGSTGRTVINTRNGDINTKGVVSAAGQVRSGAAAIANDLTKRDIY
DGAINTVTANANRSLRSGDTMTGNLTAPNFFSQNPASQPSHVPRFDQIVKDSVQDF
GYV"
915..3995
/notes="gene 37"
/codon_start=1
/transl_table=11
/protein_id="CAA24228.1"
/db_xref="PID:g15374"
/db_xref="GI:15374"
/db_xref="SWISS-PROT:P03744"
/translation="MATLKOIQFKRSKIAGTRPAASVLAEGELAINLKDRTIFTKDDSD
GNIIDGFAKGGQVQDGNVTINGLLRLNGDYVOTGGTVNGPIGSTDGVTKIFRSTQG
SFYARATNTSNAHLWFENADGTERGVIIYARPTTDDGEIRLVRROGTGTANSEFYF
RSINGGEFOANRLASDSLVTKRIADVTVIHDAKAFQYDHSVLNYYVPGTNGV
NVLKRVAKSGGTIYHEIVTAQTLADEVSWNSGDPVFKLYGRDDGRMIIRNSLAL
GFTTFNPSDGVNGVMGDKYLVLGDTVGLSYKATGTVDFLGVGGYSVASITPDSFR
STRKGI FGSEDDGATWMPGTNAALLSVOTQADNNAGDQTHIGYNAGGKNNHYFR
GTGMINITQOQMEINPGILKLVTSNNVOYFADGTISSIQPKLNDIEFLTKSNNTA
GLKFAPAGKPTDRTIQWNGGTREGQKNVYIIKAWGNSFNATGDRSRETVFQVSDSQ
GYFYAHRKAPTGTIGRIEAFQAGDVYKAGIIANGNFRVYVSSALAGNVTMSNGLF
VOGGSITGVQIGGTANALRWNAEYGAIFRSESNFYIIPNONGESGDIHSSLR
PVRIGLNDGMVGLGRDSFVQDNNALTTINSNRINANFRMLQGSAYIDAECTDAVR
PAGASFSQNNEDVRAPYMNIDRTDASAYVILKQRYVQNGGYSGLTLLNNGNFR
VYHGGGNGSTGPOTADGWEFIKNGDFISPRDLIAGRVFRDRTGNITGGSGNFANL
NSTIESLKTDMSSYPGAPIPWPSDVPAGFALMEGOTFDKSAYPKLAVAYPSGVIP
DMRGOTIKGKPSGRAVLSAEADGVKAHSHSASASTDLGKTTSSTFDYCTKGTNSTGG
HTHSGSGSTNGEHSYIEAWNGTGVGNGKMSVYIYRAGGNTNAGNHSHTFSE
GTSSAGDHSVSGIGAHHTVAIGSHGHTIITVNSGTENTVKNIAFNIVRLA"
4022..>4418
/notes="gene 38 (4418 is 1st base in codon)"
/codon_start=1
/transl_table=11
/protein_id="CAA24229.1"
/db_xref="PID:g15375"
```



```
/db_xref="GI:15375"
/db_xref="SWISS-PROT:P03739"
translation="MKIVHYFDTRFYEKYKPKVKGGLPAHSTIKKPLEPKGYA
VVFDRTODWIYEEDHRGKRAWTFNKXIFISDGSVGIIFDEPGEFDIWDGKWE
DEYKRVLIIRNRKIEELKKEQVLNMIEA" 2 others
BASE COUNT 1348 a 797 c 975 g 1296 t
ORIGIN
Query Match 15.7%; Score 176; DB 26; Length 4418;
Best Local Similarity 99.4%; Pred. No. 4.71e-119;
Matches 177; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 1 AGCTTTTCTGAAGTATCAAGAAATGGCGCATTTTCGAAACCTGCTGAATTTGGCGTCAA 60
Qy 942 asgcttttctgaagtatcaagaaatggcgcatTTTCGAAACCTGCTGAATTTGGCGTCAA 1001
Db 61 TGGTATTCGTGTAATATATCTCGCAATCCGCTTCACCTCCGGATATATGCTACTTCC 120
Qy 1002 tggattctgttaattatctcgcaatcccgcttcaccccggaataatggtacttcc 1061
Db 121 TAGCGAAGCATGCTCTAAACCTGTAAGTGTGTTGGCAAGAAATTTAGAGAAGTTTAA 178
Qy 1062 tacgcaagcatgctctaaacctgtaagtggttggcgcaagaatttagagaagtttaa 1119
RESULT 3
LOCUS MYOX2F36 902 bp DNA PHG 24-JUL-1995
DEFINITION Bacteriophage OX2 tail fiber gene 36.
ACCESSION X01753
NID G15122
VERSION X01753.1 GI:15122
KEYWORDS fiber protein; inverted repeat.
SOURCE Bacteriophage OX2
ORGANISM Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
T4-like phages.
REFERENCE 1 (bases 1 to 902)
AUTHORS Riede,I., Drexler,K. and Eschbach,M.L.
TITLE The nucleotide sequences of the tail fiber gene 36 of bacteriophage
T2 and of genes 36 of the T-even type Escherichia coli phages K3
and OX2
JOURNAL Nucleic Acids Res. 13 (2), 605-616 (1985)
MEDLINE 85215500
FEATURES
source
1..902
Location/Qualifiers
/organism="Bacteriophage OX2"
/db_xref="taxon:10691"
1..178
/gene="tail fiber gene 35"
<1..178
/gene="tail fiber gene 35"
/codon_start=2
/transl_table=11
/protein_id="CAA25890.1"
/db_xref="PID:e7867"
/db_xref="PID:g1364176"
/db_xref="GI:1364176"
/db_xref="SPTREMBL:Q38400"
translation="SFSEVSRNGISKPAEFGVNGIRVNYVCESAPDPIMVLPTQAS
SKTGKVFQGEFEV"
184..201
/misc_signal /note="put. transcription terminator"
184..190
/repeat_unit /note="repeat A"
/rpt_type=INVERTED
195..201
/repeat_unit /note="repeat A"
/rpt_type=INVERTED
210..222
/promoter /note="putative jukebox for late transcription initiation
(g36,g37,g38)"
229..232
/RBS /note="Shine Dalgarno sequence"
```

```
gene
241..891
/gene="tail fiber gene 36 (aa 1-216)"
CDS
241..891
/gene="tail fiber gene 36 (aa 1-216)"
/codon_start=1
/transl_table=11
/protein_id="CAA25891.1"
/db_xref="PID:g15123"
/db_xref="GI:15123"
/db_xref="SWISS-PROT:P07627"
translation="MADLKVGSTVGGSVIWHQGNFPLNSAGDDVLYKFKIYSEYNKP
QAANDFYVSKANGTGTGPTITINYGVSYLQLSNNETPIRISGGGTGNTLVGGSSG
GISFRPAGSEITTGQITITPEGLTFTTRAVTAPSTITVTPSAADVTVRKDVDCGAIN
TVTNANSRVLRSBGDTMTGNLTAPNLFSONPASQPSHPVRFDOIVIKDSVQDFGY"
900..902
/gene="tail fiber gene 37"
900..>902
/gene="tail fiber gene 37"
/codon_start=1
/transl_table=11
/protein_id="CAA25892.1"
/db_xref="PID:e7869"
/db_xref="PID:g4379274"
/db_xref="GI:4379274"
translation="M"
BASE COUNT 273 a 176 c 176 g 277 t
ORIGIN
Query Match 13.9%; Score 156; DB 26; Length 902;
Best Local Similarity 93.8%; Pred. No. 4.83e-102;
Matches 167; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
Db 1 AAGCTTTTCTGAAGTATCAAGAAATGGCGCATTTTCGAAACCTGCTGAATTTGGCGTCAA 60
Qy 942 asgcttttctgaagtatcaagaaatggcgcatTTTCGAAACCTGCTGAATTTGGCGTCAA 1001
Db 61 CGGTATTCGCTTAATATGCTGCAATCGGCTTCACCTCCAGATATATGCTACTTCC 120
Qy 1002 tggattctgttaattatctatctcgcaatcccgcttcaccccggaataatggtacttcc 1061
Db 121 TACCAAGCCTCTCTTAAACCTGGCAAGTGTGTTGGCAAGAAATTTAGAGAAGTTTAA 178
Qy 1062 tacgcaagcatgctctaaacctgtaagtggttggcgcaagaatttagagaagtttaa 1119
RESULT 4
LOCUS MY2F36 911 bp DNA PHG 17-FEB-1997
DEFINITION Bacteriophage T2 tail fiber gene 36.
ACCESSION X01755
NID G15189
VERSION X01755.1 GI:15189
KEYWORDS fiber protein; inverted repeat.
SOURCE coliphage T2
ORGANISM Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
T4-like phages.
1 (bases 1 to 911)
AUTHORS Riede,I., Drexler,K. and Eschbach,M.L.
TITLE The nucleotide sequences of the tail fiber gene 36 of bacteriophage
T2 and of genes 36 of the T-even type Escherichia coli phages K3
and OX2
JOURNAL Nucleic Acids Res. 13 (2), 605-616 (1985)
MEDLINE 85215500
REFERENCE 2 (bases )
AUTHORS Riede,I.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-1985) to the EMBL/GenBank/DBSJ databases
COMMENT Data kindly reviewed (18-Nov-1985) by Riede I.
FEATURES
source
1..911
Location/Qualifiers
/organism="coliphage T2"
/db_xref="taxon:10664"
1..178
gene
```

```
/gene="tail fiber gene 35"
<1..178
/gene="tail fiber gene 35"
/codon_start=2
/transl_table=11
/protein_id="CAA25896.1"
/db_xref="PID:g15190"
/db_xref="GI:15190"
/db_xref="SPTREMBL:Q38425"
/translation="SFSEVSRNGSISKPAEFGVNGIRVNVYCESASPPDIIVLPTQAS
SKTGKVFQGEFREV"
misc_signal
184..201
/notes="put. transcription terminator"
repeat_unit
184..190
/notes="imp. repeat A"
/pt_type="INVERTED"
repeat_unit
195..201
/notes="imp. repeat A"
/pt_type="INVERTED"
promoter
210..222
/notes="put."
RBS
229..232
/notes="rRNA binding site"
gene
241..900
/gene="tail fiber gene 36 (aa 1-219)"
CDS
241..900
/gene="tail fiber gene 36 (aa 1-219)"
/transl_table=11
/protein_id="CAA25897.1"
/db_xref="PID:g15191"
/db_xref="GI:15191"
/db_xref="SPTREMBL:Q38426"
/translation="MADLVKVGSTGGSVIWHQGNFPLNPGADDDVLYKSKFIYSEYNKP
QAADNVLKANGTGLNVRIGNQVPGAGNLTGMVPGDGTATNDVNLDIISW
YGIKSSOGTGPRTIVNVRGEISARGNINSORQVRAEAPAAATNDLTRKIVDGG
AINTVITANANRSLRSGDTNLTAPNFSSQSPSHVPRFDQIVIKDSVQDFGY"
old_sequence
720
/gene="tail fiber gene 36 (aa 1-219)"
/citation={1}
/replace="t"
gene
909..911
/gene="tail fiber gene 37"
CDS
909..911
/gene="tail fiber gene 37"
/codon_start=1
/transl_table=11
/protein_id="CAA25898.1"
/db_xref="PID:e7916"
/db_xref="PID:g4379275"
/db_xref="GI:4379275"
/translation="M"
BASE COUNT 269 a 171 c 200 g 271 t
ORIGIN

Query Watch 13.6%; Score 152; DB 26; Length 911;
Best Local Similarity 92.7%; Pred. No. 1.17e-98;
Matches 165; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Db 1 AACCTTTCTGAGGTATCAAGAAATGGTAGTATTTCGAACCCGCTGAATTTGGTGTCAA 60
QY 942 aagctttctgaagtatcaagaataatgcgcatttcgaacacctgcgaattggcgtcaa 1001

Db 61 TGGTATTCGAGTTAAATATGCTGTGATCTGCTTCACCTCCGGATATATGCTACTTC 120
QY 1002 tggattctggttaatatatctatctgcgaatccgcttcaccccggaataatggacttcc 1061

Db 121 TACACAAGCATCGCTAAACATGGTAAAGCTGTTTGGCGCAAGATTTAGAGAGATATAA 178
QY 1062 tacgcaagcatcgctctaaactggttaagtgtttggcgcaagaatttagagaagttaa 1119
```

```
5
LOCUS MYK3F36 944 bp DNA PHG 24-JUL-1995
DEFINITION Bacteriophage K3 tail fiber gene 36.
ACCESSION X01754
NID 915108
VERSION X01754.1 GI:15108
KEYWORDS fiber protein; inverted repeat.
SOURCE Bacteriophage K3.
ORGANISM Bacteriophage K3.
REFERENCE 1 (bases 1 to 944)
AUTHORS Riede,I., Drexler,K. and Eschbach,M.L.
TITLE The nucleotide sequences of the tail fiber gene 36 of bacteriophage
T2 and of genes 36 of the T-even type Escherichia coli phages K3
and OX2
JOURNAL Nucleic Acids Res. 13 (2), 605-616 (1985)
MEDLINE 85215500
FEATURES
source
Location/Qualifiers
1..944
/organism="Bacteriophage K3"
/db_xref="taxon:10674"
1..178
/gene="tail fiber gene 35"
<1..178
/codon_start=2
/gene="tail fiber gene 35"
/transl_table=11
/protein_id="CAA25893.1"
/db_xref="PID:e7855"
/db_xref="PID:g1364175"
/db_xref="GI:1364175"
/db_xref="SPTREMBL:Q38393"
/translation="SFSEVSRNGSISKPAEFGVNGIRVNVYCESASPPDIIVLPTQAS
SKTGKVFQGEFREV"
misc_signal
184..201
/notes="put. transcription terminator"
repeat_unit
184..190
/notes="repeat A"
/pt_type="INVERTED"
repeat_unit
195..201
/notes="repeat A"
/pt_type="INVERTED"
promoter
210..222
/notes="putative jukebox for late transcription initiation
(936,937,938)"
RBS
229..232
/notes="Shine Dalgarno sequence"
241..933
/gene="tail fiber gene 36 (aa 1-230)"
241..933
/gene="tail fiber gene 36 (aa 1-230)"
/codon_start=1
/transl_table=11
/protein_id="CAA25894.1"
/db_xref="PID:g15109"
/db_xref="GI:15109"
/db_xref="SPTREMBL:Q38425"
/translation="MADLVKVGSTGGSVIWHQGNFPLNPGADDDVLYKSKFIYSEYNKP
QAADNVLKANGTGLNVRIGNQVPGAGNLTGMVPGDGTATNDVNLDIISW
YGIKSSOGTGPRTIVNVRGEISARGNINSORQVRAEAPAAATNDLTRKIVDGG
AINTVITANANRSLRSGDTNLTAPNFSSQSPSHVPRFDQIVIKDSVQDFGY"
942..944
/gene="tail fiber gene 37"
942..944
/gene="tail fiber gene 37"
/codon_start=1
/transl_table=11
/protein_id="CAA25895.1"
/db_xref="PID:e7857"
/db_xref="PID:g4379273"
/db_xref="GI:4379273"
/translation="M"
```

[illegible]



[illegible]

```

/organism="Oxytricha fallax"
/strain="9D1"
/db_xref="taxon:5944"
/transposon="TBE1"
/note="this is a bulk sequence that was generated from a
PCR product that represents many transposon templates"
<1. >354
/codon_start=1
/transl_table=6
/product="57kD zinc finger/protein chimera"
/protein_id="AAB49644.1"
/db_xref="PID:g181676"
/db_xref="GI:181676"
/translation="HTRDLXKHLKKAHKXXEXEXXXLKLXKRKAREXXXXXXXXX
QAXEXXVXXRXNLXASEXTKIMIKIQYKKIPVLAQIDLDTSLSYLILEDSDFKKVI
XDHOYAYKVVPENL"
BASE COUNT      106 a      42 c      41 g      54 t      111 others
ORIGIN

Query Match      3.0% Score 34; DB 21; Length 354;
Best Local Similarity 26.5%; Pred. No. 1.13e-05;
Matches 36; Conservative 49; Mismatches 51; Indels 0; Gaps 0;

Db 41 ARAAGADWTAGARRTTCARRAKHAWCAMPHWARYTRAAGYTDCTCAHAAGAAGAARG 100
Cp 765 agatattgttagttccagagcaagtaaatcagcatcagcataaactgagctttaag 706
Db 101 CDAGRGARYHHYHDWAGWBGYDQWGYKAAAYAAAGCAGWAGRWYASWTRYGTAAARGA 160
Cp 705 ataaagaacatctccaggaattaataataatcagataattacttatagcagctgaagt 646
Db 161 WDKRCGCAACCCWYT 176
Cp 645 tggcaatctcgcaatt 630

RESULT 11
LOCUS      128278      215 bp      DNA      PAT      30-OCT-1996
DEFINITION      Sequence 5 from patent US 5569830.
ACCESSION      128278
NID      g1819054
KEYWORDS      128278.1 GI:1819054
SOURCE      Unknown.
ORGANISM      Unknown.
REFERENCE      1 (bases 1 to 215)
AUTHORS      Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
TITLE      Plant inhibitors of fungal polygalacturonases and their use to
control fungal disease
JOURNAL      Patent: US 5569830-A 5 29-OCT-1996;
FEATURES      Location/Qualifiers
source      1..215
            /organism="unknown"
BASE COUNT      15 a      8 c      25 g      26 t      141 others
ORIGIN

Query Match      2.9% Score 32; DB 25; Length 215;
Best Local Similarity 17.2%; Pred. No. 2.01e-04;
Matches 28; Conservative 63; Mismatches 70; Indels 2; Gaps 2;

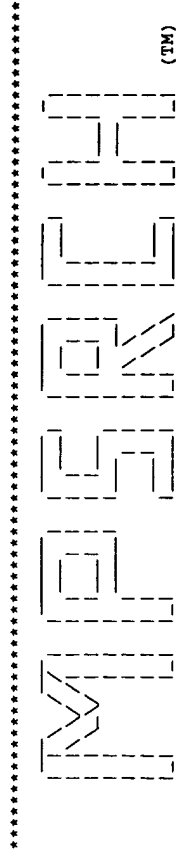
Db 20 DKAKKDGNTTSSMTDCCNRTGWCVCDTDTT-YRVNNDSGHNKYSSANYVYGGNVGAAKT 78
QY 520 ttgagatgttatttgacagcttagaagatgtagggctaccgggtttccaagaggaac 579
Db 79 HYTHYVNSGADSKTYTDSYNASGTTSSNGTDCGNSGADSYGSSKTAAMTSRRRTGKTAN 138
QY 580 tatgaaagtgttagcaaatcatcgtcggaagtgttggaactaatg-acgaaattcgag 638
Db 139 NAVDSRNMGDASVGSKNTKKKHAKNSADGKVGSKNNGDRNNRY 181
QY 639 attgcaacttcagctgataagataaattatctgattataat 681

```

RESULT 12  
LOCUS OFU89262 370 bp DNA INV 14-MAR-1997  
DEFINITION Oxytricha fallax 57kd zinc finger/protein kinase gene, partial cds.  
ACCESSION U89262  
NID 91881681  
VERSION U89262.1 GI:1881681  
KEYWORDS  
SOURCE  
ORGANISM Oxytricha fallax.  
Eukaryotae; mitochondrial eukaryotes; Alveolata; Ciliophora;  
hypotrichs; Stichotrichida; Oxytricha.  
REFERENCE 1 (bases 1 to 370)  
AUTHORS Doak,T.G., Doerder,F.P., Jahn,C.L. and Herrick,G.  
TITLE A proposed superfamily of transposase genes: transposon-like  
elements in ciliated protozoa and a common 'D35E' motif  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (3), 942-946 (1994)  
MEDLINE 9413477  
REFERENCE 2 (bases 1 to 370)  
AUTHORS Witherspoon,D.J., Doak,T.G., Williams,K., Seger,J. and Herrick,G.  
TITLE Selection on the protein-coding genes of the TB1 family of  
transposable elements in the ciliates Oxytricha fallax and O.  
trifallax  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 370)  
AUTHORS Doak,T.G., Williams,K., Witherspoon,D.J. and Herrick,G.  
TITLE Direct Submission  
JOURNAL Submitted (11-FEB-1997) Oncological Science, University of Utah,  
School of Med. Rm5C334, USA, UT 84132, USA  
FEATURES  
source Location/Qualifiers  
1..370  
/organism="Oxytricha fallax"  
/strain="9D1"  
/db\_xref="taxon:5944"  
/transposon="TB1"  
/note="this is a bulk sequence that was generated from a  
PCR product that represents many transposon templates"  
<1..>370  
/codon\_start=3  
/transl\_table=6  
/product="57kd zinc finger/protein kinase"  
/protein\_id="AAB49647.1"  
/db\_xref="PID:g1881682"  
/db\_xref="GI:1881682"  
/translation="HPRDLXKLLKAKKXXEXXXXXXXLXNKKRAREXXXXXX  
OAXEYVXXRXNLXSEXTKIMIKIYKIPVLAQIDLTSLQSLILEDSFDKVI  
XDHQYAYKVPENYLSDDK"  
BASE COUNT 116 a 43 c 48 g 69 t 94 others  
ORIGIN  
Query Match 2.98; Score 33; DB 21; Length 370;  
Best Local Similarity 33.18; Pred. No. 4.81e-05;  
Matches 45; Conservative 39; Mismatches 52; Indels 0; Gaps 0;  
Db 43 ARAAGANWTAGARRTTGARRAKHAWGAMRHTSARYTGAGYDCTMAACAAGAAAG 102  
Cp 765 agatatttctgttccagagcaagtaaatcagcatcagcataaactgagcttaag 706  
Db 103 CRAGRGARYMTYDWARGYBGWGWGCTTKAAYAAGCWAGHAGTAYGAGTAYGTDARGA 162  
Cp 705 ataaagaacatctccaggaattaaattataatcagataattacttattatagcagctgaagt 646  
Db 163 WDKACGCAACCCWMT 178  
Cp 645 tggcaatctcgcaatt 630  
RESULT 13  
LOCUS AR024229 965 bp DNA PAT 04-DEC-1998  
DEFINITION Sequence 22 from patent US 5795961.  
ACCESSION AR024229  
NID 93977523  
VERSION AR024229.1 GI:3977523

KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 965)  
AUTHORS Wallace,T.Paul, Harris,W.J., Carr,F.J., Old,L.J., Welt,S. and  
Kitamura,K.  
TITLE Recombinant human anti-Lewis b antibodies  
JOURNAL Patent: US 5795961-A 22 18-AUG-1998;  
FEATURES Location/Qualifiers  
source 1..965  
BASE COUNT 192 a 170 c 226 g 205 t 172 others  
ORIGIN  
Query Match 2.98; Score 33; DB 25; Length 965;  
Best Local Similarity 22.88; Pred. No. 4.81e-05;  
Matches 21; Conservative 45; Mismatches 24; Indels 2; Gaps 2;  
Db 812 RGRGWGDYGGYTNNGKRGRTVMTDTSNSRSSVTA-ADTAVYVCVRGRSDSDGGDYW 870  
QY 823 gggcaagctgggctcatgggaattaaaggaattatgtagttgtccagaaaacgagta 882  
Db 871 GGTVTVSSHTVKDMTSSSSASVGDRTVTCRS 902  
QY 883 ggattacgatatat-gcacagagaactgcac 913  
RESULT 14  
LOCUS HUAC004787 216021 bp DNA PRI 24-JUL-1998  
DEFINITION Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10, complete  
sequence.  
ACCESSION AC004787  
NID g3337381  
VERSION AC004787.1 GI:3337381  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 216021)  
AUTHORS Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,  
Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.  
TITLE Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 216021)  
AUTHORS Adams,M.D. and Loftus,B.J.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1998) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA, Email:  
b.loftus@tigr.org  
REFERENCE 3 (bases 1 to 216021)  
AUTHORS Adams,M.D. and Loftus,B.J.  
TITLE Direct Submission  
JOURNAL Submitted (24-JUL-1998) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
COMMENT On Jul 24, 1998 this sequence version replaced gi:3241936.  
Address all correspondence to: Mark Adams The Institute for Genomic  
Research 9712 Medical Center Dr., Rockville, MD 20850, USA e-mail  
address: humgen@tigr.org. The orientation of the sequence is from  
SP6 to T7 end. Genes were identified by a combination of five  
methods including: XGRAIL (available by anonymous ftp from  
arthur.epm.ornl.gov), Genefinder (Phil Green, University of  
Washington), Genscan (Chris Burge,  
http://genomic.stanford.edu/~chris/GENSCANW.html) searches of the  
complete sequence against a peptide database, and the Human gene  
Index database at TIGR (http://www.tigr.org/tdb/hgi.html).  
Genes without peptide homology having spliced EST hits are termed  
'Unknown gene product'. Genes encoding tRNAs are predicted by  
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/CRNscan-SE/).  
FEATURES Location/Qualifiers  
source 1..216021  
/organism="Homo sapiens"





Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:00:35 1999; MasPar time 13.02 Seconds  
Tabular output not generated. 807.364 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (1-372) from PCTUS9913024.pep (1 of 12)  
Perfect Score: 2542  
Sequence: 1 MEKFAEFGQGVQTPFLSE.....LPTQASSKTKGVFGQEPREV 372

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 49.734; Variance 92.468; scale 0.538

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1893	74.5	295	1	TAIL FIBER PROTEIN GP3	0.00e+00
2	300	11.8	65	1	HYPOTHETICAL 7.3 KD PR	3.10e-35
3	124	4.9	196	1	ALKH_BACSU	7.12e-04
4	113	4.4	323	1	HEM2_HELPY	2.56e-02
5	108	4.2	672	1	NOC3_PARDE	1.21e-01
6	106	4.2	673	1	YAZ1_SYNY3	2.23e-01
7	106	4.2	2492	1	POLN_EEVT	2.23e-01
8	103	4.1	154	1	OM22_NEUCR	5.46e-01
9	103	4.1	390	1	DHE4_SULSH	5.46e-01
10	103	4.1	723	1	REFX2_HUMAN	5.46e-01
11	105	4.1	727	1	NUAM_BOVIN	3.01e-01
12	105	4.1	727	1	NUAM_HUMAN	3.01e-01
13	105	4.1	1500	1	SSP5_STRGN	3.01e-01
14	104	4.1	2492	1	POLN_EEVV3	4.06e-01
15	99	3.9	421	1	DHE3_SULSO	1.75e+00
16	99	3.9	475	1	NCAP_SYN	1.75e+00
17	98	3.9	699	1	YAF6_SCHPO	2.32e+00
18	100	3.9	861	1	MUTS_HAEIN	1.31e+00
19	99	3.9	945	1	AMPE_MOUSE	1.75e+00
20	100	3.9	1300	1	POL2_MOUSE	1.31e+00
21	100	3.9	1693	1	POLN_HEVNY	1.31e+00
22	99	3.9	2492	1	POLN_EEVP	1.75e+00
23	97	3.8	380	1	CAPM_STAAU	3.08e+00

24	97	3.8	542	1	OPPA_SALTY	PERIPLASMIC OLIGOPEPTI	3.08e-00
25	97	3.8	578	1	SHC_MOUSE	SHC TRANSFORMING PROTE	3.08e-00
26	97	3.8	583	1	SHC_HUMAN	SHC TRANSFORMING PROTE	3.08e-00
27	96	3.8	692	1	REFX2_MOUSE	DNA BINDING PROTEIN RF	4.08e-00
28	97	3.8	780	1	ACON_BOVIN	ACONITATE HYDRATASE, M	3.08e-00
29	96	3.8	833	1	GYRA_BACS5	DNA GYRASE SUBUNIT A (	4.08e-00
30	96	3.8	969	1	MSUI_YEAST	MITOCHONDRIAL BIOTENES	4.08e-00
31	93	3.7	156	1	YEHS_ECOLI	HYPOTHETICAL 18.0 KD P	9.31e+00
32	95	3.7	171	1	YAL4_PHVU	HYPOTHETICAL AL4 PROTE	5.39e-00
33	95	3.7	196	1	YCH6_YEAST	HYPOTHETICAL 22.3 KD P	5.39e-00
34	94	3.7	303	1	ILID_ECOLI	HYPOTHETICAL ABC TRANS	7.09e-00
35	95	3.7	334	1	SYW_HAEIN	TRYPHOPHANYL-TRNA SYN	5.39e-00
36	95	3.7	393	1	FLAA_PSEAE	FLAGELLIN.	5.39e-00
37	95	3.7	432	1	TIG_ECOLI	TRIGGER FACTOR (TF).	5.39e-00
38	95	3.7	499	1	NU4C_MARPO	NADH-PLASTOQUINONE OXI	5.39e-00
39	94	3.7	577	1	URE1_MYCTU	UREASE ALPHA SUBUNIT (	7.09e-00
40	94	3.7	663	1	GLI3_CHICK	ZINC FINGER PROTEIN GL	7.09e-00
41	94	3.7	776	1	VP42_ROTTS	OUTER CAPSID PROTEIN V	7.09e-00
42	94	3.7	960	1	KKIT_CHICK	MAST/STEM CELL GROWTH	7.09e-00
43	94	3.7	997	1	YPX2_CAEEL	HYPOTHETICAL 110.3 KD	7.09e-00
44	94	3.7	1354	1	VGR2_HUMAN	VASCULAR ENDOTHELIAL G	7.09e-00
45	95	3.7	1958	1	UBRI_SCHPO	PROBABLE N-END-RECOGNI	5.39e-00

ALIGNMENTS

RESULT 1  
ID VG35\_BPT4 STANDARD; PRT; 295 AA.  
AC P03742;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE TAIL FIBER PROTEIN GP35.  
GN 25.  
OS BACTERIOPHAGE T4.  
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE;  
OC T4-LIKE PHAGES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SELIVANOF N., REVEL H., MESYANZHINOV V.V., GOLDBER E.;  
RL UNPUBLISHED OBSERVATIONS (XXX-1991).  
RN [2]  
RP SEQUENCE OF 238-295 FROM N.A.  
RX MEDLINE: 82170495.  
RA OLIVER D.B., CROWTHER R.A.;  
RT "DNA sequence of the tail fibre genes 36 and 37 of bacteriophage T4.";  
RL J. MOL. BIOL. 153:545-568(1981).  
CC [1]-FUNCTION: STRUCTURAL COMPONENT OF THE DISTAL-HALF TAIL FIBER.  
CC THE TAIL FIBER OF T4 IS ABOUT 1600 ANGSTROMS LONG WITH A KINK IN  
CC THE MIDDLE THAT DIVIDES THE FIBER INTO PROXIMAL AND DISTAL HALVES.  
CC THE THIN TIP OF THE DISTAL HALF-FIBER INTERACTS WITH THE BACTERIAL  
CC LIPOPOLYSACCHARIDE RECEPTOR AND SPECIFIES THE HOST RANGE OF THE  
CC PHAGE.  
CC [1]-SUBUNIT: THE DISTAL HALF-FIBER CONTAINS TWO MOLECULES EACH OF GP36  
CC AND GP37 AND ONE MOLECULE OF GP35.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: V00863; G836604; -;  
CC EMBL: J02509; G215885; -;  
CC DR PIR: A04367; TLBP54.  
CC DR PIR: J0577; JT0577.  
KW FIBER PROTEIN  
SQ SEQUENCE 295 AA; 32282 MW; 2F9ACC4A CRC32;

Query Match 74.5%; Score 1893; DB 1; Length 295;  
Best Local Similarity 99.3%; Pred. No. 0.00e+00;



```
Matches 277; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 17 FMNSLTNNRIVAITLTSKVNPPPEVSWLRTAGTSAPSDSILSRFDVSYAAPTSSKRA 76
Qy 94 FMNSLTNNRIVAITLTSKVNPPPEVSWLRTAGTSAPSDSILSRFDVSYAAPTSSKRA 153
Db 77 IALSHVLSNRKSTDDYQOTILDVVFDSLEDVYGATGFPRTYSEVQFMSAVGGTNNIEAR 136
Qy 154 IALSHVLSNRKSTDDYQOTILDVVFDSLEDVYGATGFPRTYSEVQFMSAVGGTNDIEAR 213
Db 137 LPTSAISKSLDYNLIPGDVLYLKAQLYADADLLALGTNNISIRFYNASNGYISSTQAEF 196
Qy 214 LPTSAISKSLDYNLIPGDVLYLKAQLYADADLLALGTNNISIRFYNASNGYISSTQAEF 273
Db 197 TGOAGSWELKEDYVVPVENAVGFTIYAQRTAAGOGGMRNLFSFSEVSRNGGISKPAEFGV 256
Qy 274 TGOAGSWELKEDYVVPVENAVGFTIYAQRTAAGOGGMRNLFSFSEVSRNGGISKPAEFGV 333
Db 257 NGRVNVICESASPPDIWLPQTQASSKTGKVGQEFREV 295
Qy 334 NGRVNVICESASPPDIWLPQTQASSKTGKVGQEFREV 372
RESULT 2
ID Y15A_BP74 STANDARD; PRT; 65 AA.
AC P39509;
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DE HYPOTHETICAL 7.3 KD PROTEIN IN GP34-GP35 INTERGENIC REGION.
GN Y15A OR 34.1.
OS BACTERIOPHAGE T4.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE;
OC T4-LIKE PHAGES.
RN [1]
RA KUTTER E.M.;
RL SUBMITTED (NOV-1994) TO THE SWISS-PROT DATA BANK.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 65 AA; 7334 MW; C2D7CE2D CRC32;
Query Match 11.8%; Score 300; DB 1; Length 65;
Best Local Similarity 82.1%; Pred. No. 3.10e-35;
Matches 46; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
Db 1 MEKMAEITWTRICPNAILSESRYKISITAGSCPLSTAGPSYVKFQDNVPGSQTF 56
Qy 1 MEKMAEITWTRICPNAILSESRYKISITAGSCPLSTAGPSYVKFQDNVPGSQTF 56
RESULT 3
ID ALKH_BACSU STANDARD; PRT; 196 AA.
AC P50846;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DE 4-HYDROXY-2-OXOGLUTARATE ALDOLASE (EC 4.1.3.16) (2-KETO-4-
DE HYDROXYGLUTARATE ALDOLASE) (KHG-ALDOLASE) / 2-DEHYDRO-3-
DE DEOXYPHOSPHOGLUCONATE ALDOLASE (EC 4.1.2.14) (PHOSPHO-2-DEHYDRO-3-
DE DEOXYGLUCONATE ALDOLASE) (PHOSPHO-2-KETO-3-DEOXYGLUCONATE ALDOLASE)
DE (2-KETO-3-DEOXY-6-PHOSPHOGLUCONATE ALDOLASE) (KDPG-ALDOLASE).
GN KDA.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RA SEQUENCE FROM N.A.
RC STRAIN-168 / MARBURG;
RX MEDLINE; 96349105.
RA SOROKIN A.V., AZEVEDO V., ZUMSTEIN E., GALLERON N., EHRlich S.D.,
RA SERRON P.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
RT the serA and kdg loci cloned in a yeast artificial chromosome.";
```

```
RL MICROBIOLOGY 142:2005-2016(1996).
CC -1- CATALYTIC ACTIVITY: 4-HYDROXY-2-OXOGLUTARATE - PYRUVATE +
CC GLYOXYLATE.
CC -1- CATALYTIC ACTIVITY: 2-DEHYDRO-3-DEOXY-D-GLUCONATE 6-PHOSPHATE -
CC PYRUVATE + D-GLYCERALDEHYDE 3-PHOSPHATE.
CC -1- PATHWAY: KEY ENZYME IN THE ENTNER-DOUDOROFF PATHWAY.
CC -1- PATHWAY: PARTICIPATES IN THE REGULATION OF THE INTRACELLULAR LEVEL
CC OF GLYOXYLATE.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE KHG/KDPG ALDOLASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L47838; G1146190; -
CC EMBL; Z99115; E1183656; -
CC DR SUBTILIS; BG11396; KDA.
CC DR PROSITE; PS00159; ALDOLASE_KDPG_KHG_1; FALSE_NEG.
CC DR PROSITE; PS00160; ALDOLASE_KDPG_KHG_2; 1.
CC DR PFAM; PF01081; Aldolase; 1.
CC KW LYASE; SCHIFF BASE; MULTIFUNCTIONAL ENZYME.
CC FT ACT_SITE 43 43 BY SIMILARITY.
CC FT ACT_SITE 47 47 BY SIMILARITY.
CC FT BINDING 130 130 SCHIFF-BASE WITH KHG OR PYRUVATE (BY
CC SIMILARITY).
CC SQ SEQUENCE 196 AA; 20865 MW; 14FCBED5 CRC32;
Query Match 4.9%; Score 124; DB 1; Length 196;
Best Local Similarity 27.8%; Pred. No. 7.12e-04;
Matches 20; Conservative 27; Mismatches 21; Indels 4; Gaps 4;
Db 114 TPSEIMEALTFGTTLKLFPSGVGIFPMKNLAGPFPQVTFIPFGGIIH-PSEVPDWLR-A 171
Qy 68 STGALVDSKSIATFSTNDTTSAAF-VSFMSNLNN-RIVAILTSGKVNFPPEVVSRLTA 125
Db 172 GAGAVGVGSQLG 183
Qy 126 GTSAPFSDSILS 137
RESULT 4
ID HEM2_HELPY STANDARD; PRT; 323 AA.
AC P56074;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE DELTA-AMINOLEVULINIC ACID DEHYDRATASE (EC 4.2.1.24) (PORPHOBILINOGEN
DE SYNTHASE) (ALADH).
GN HEMB OR HP0163.
OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; HELICOBACTER GROUP;
OC HELICOBACTER.
RN [1]
RA SEQUENCE FROM N.A.
RC STRAIN-26695 / ATCC 700392;
RX MEDLINE; 97394467.
RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLOCKEY A.,
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATHEY L., WALLIN E.,
RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA VENTER J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
```

RL NATURE 388:539-547(1997).  
CC -1- CATALYTIC ACTIVITY: 2 5-AMINOLEVULINATE - PORPHOBILINOGEN +  
CC 2 H(2)O.  
CC -1- COFACTOR: ZINC.  
CC -1- PATHWAY: SECOND STEP IN PORPHYRIN BIOSYNTHESIS.  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF000537; G2313250; -  
CC TIGR; HP0163; -  
CC PROSITE; PS00169; D\_ALA\_DEHYDRATASE; 1.  
CC PFAM; PF00490; ALAD; 1.  
CC PORPHYRIN BIOSYNTHESIS; LYASE; ZINC.  
CC KW DOMAIN 115 132 ZINC-BINDING (BY SIMILARITY).  
CC FT ACT\_SITE 246 246 BY SIMILARITY.  
CC SQ SEQUENCE 323 AA; 36202 MW; 9B60D720 CRC32;  
Query Match 4.4%; Score 113; DB 1; Length 323;  
Best Local Similarity 25.7%; Pred. No. 2,56e-02;  
Matches 27; Conservative 33; Mismatches 39; Indels 6; Gaps 6;  
Db 135 SVSNKDT-LEILNLOGLLAESGVDLAPSNM-MDCNVLS-LRKTLDNAGYHTPTMS-Y 190  
QY 81 STSNDTTSAAFFSMNSLNNRVAITSGKVNFPPEVSWLRTA-GTSAFFSDSILSRF 139  
Db 191 STKFASYGPPFDVANSAPSGFGRKSYQMDYANOKALELSELD 235  
QY 140 DVSYAAYFTSSKRAIALEHVKLNSRKSTD-DYQTILDVDFDSLED 183  
RESULT 5  
ID NQ03\_PARDE STANDARD; PRT; 672 AA.  
AC P29915;  
DT 01-APR-1993 (REL. 25, CREATED)  
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 3 (EC 1.6.5.3) (NADH  
DE DEHYDROGENASE 1, CHAIN 3) (NDH-1, CHAIN 3).  
GN NQ03.  
OS PARACOCUS DENITRIFICANS.  
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHODOBACTER GROUP;  
OC PARACOCUS.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.  
RC STRAIN-ATCC 13548;  
RX MEDLINE; 92296779.  
RA XU X., MATSUNO-YAGI A., YAGI T.;  
RT "Structural features of the 66-kDa subunit of the energy-transducing  
RT NADH-ubiquinone oxidoreductase (NDH-1) of Paracoccus denitrificans.";  
RL ARCH. BIOCHEM. BIOPHYS. 296:40-48(1992).  
RN [2]  
RP SEQUENCE OF 657-672 FROM N.A.  
RX MEDLINE; 93136200.  
RA XU X., MATSUNO-YAGI A., YAGI T.;  
RT "DNA sequencing of the seven remaining structural genes of the gene  
RT cluster encoding the energy-transducing NADH-quinone oxidoreductase  
RT of Paracoccus denitrificans.";  
RL BIOCHEMISTRY 32:968-981(1993).  
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE - NAD(+) + UBIQUINOL.  
CC -1- COFACTOR: MAY BIND TWO 4FE-4S CLUSTER AND ONE 2FE-2S CLUSTER.  
CC -1- SUBUNIT: COMPOSED OF 14 DIFFERENT SUBUNITS.  
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 75 KD SUBUNIT FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D90901; G1652020; -  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M84572; G150601; -  
CC PIR; S23948; S23948.  
CC PROSITE; PS00641; COMPLEX1\_75K\_1; 1.  
CC PROSITE; PS00642; COMPLEX1\_75K\_2; 1.  
CC PROSITE; PS00643; COMPLEX1\_75K\_3; 1.  
CC PFAM; PF00111; fer2; 1.  
CC PFAM; PF00384; molybdopterln; 1.  
CC KW OXIDOREDUCTASE; NAD; UBIQUINONE; IRON-SULFUR; 4FE-4S.  
CC FT INIT\_MET 0  
CC METAL 25 25 IRON-SULFUR (2FE-2S) (POTENTIAL).  
CC METAL 36 36 IRON-SULFUR (2FE-2S) (POTENTIAL).  
CC METAL 47 47 IRON-SULFUR (2FE-2S) (POTENTIAL).  
CC METAL 50 50 IRON-SULFUR (2FE-2S) (POTENTIAL).  
CC METAL 109 109 IRON-SULFUR (2FE-2S) (POTENTIAL).  
CC METAL 112 112 IRON-SULFUR (2FE-2S) (POTENTIAL).  
CC METAL 118 118 IRON-SULFUR (2FE-2S) (POTENTIAL).  
CC METAL 137 137 IRON-SULFUR (2FE-2S) (POTENTIAL).  
CC METAL 160 160 IRON-SULFUR (4FE-4S) (POTENTIAL).  
CC METAL 163 163 IRON-SULFUR (4FE-4S) (POTENTIAL).  
CC METAL 207 207 IRON-SULFUR (4FE-4S) (POTENTIAL).  
CC SQ SEQUENCE 672 AA; 73028 MW; 40DB761A CRC32;  
Query Match 4.2%; Score 108; DB 1; Length 672;  
Best Local Similarity 26.5%; Pred. No. 1,21e-01;  
Matches 22; Conservative 26; Mismatches 30; Indels 5; Gaps 5;  
Db 164 VRFTTEVAGITOMGTGRG-EDSEITSYLNQTLSENMGNIDLCVPGLV-SKPYAFTA 221  
QY 24 VRKISIAQCLSTAGPSYVVFQDNPVGSQTFSAGLHLRVFDP-STGALVDKSKYAFST 82  
Db 222 RPWELTKTESIDVMDALGSSIRI 244  
QY 83 SN-DTTSAAFFSMNSLTNN-RI 103  
RESULT 6  
ID YA21\_SYNY3 STANDARD; PRT; 673 AA.  
AC P72929;  
DT 15-DEC-1998 (REL. 37, CREATED)  
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 74.4 KD PROTEIN SLL1021.  
GN SLL1021.  
OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).  
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97061201.  
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,  
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,  
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,  
RA TABATA S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions.";  
RL DNA RES. 3:109-136(1996).  
CC -1- SIMILARITY: TO B.SUBTILIS YUAG AND E.COLI YQIK.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D90901; G1652020; -





```
DR PIR: A33552; A33552.
DR PROSITE; PS00641; COMPLEX1_75K_1; 1.
DR PROSITE; PS00642; COMPLEX1_75K_2; 1.
DR PROSITE; PS00643; COMPLEX1_75K_3; 1.
DR PFAM; PF00111; fer2; 1.
DR PFAM; PF00384; molybdopterin; 1.
KW OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION; TRANSIT PEPTIDE;
KW IRON-SULFUR; 4FE-4S.
FT TRANSIT 1 23 MITOCHONDRION.
FT CHAIN 24 727 NADH-UBIQUINONE REDUCTASE 75 KD SUBUNIT.
FT METAL 53 53 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 64 64 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 75 75 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 78 78 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 92 92 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 128 128 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 131 131 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 137 137 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 176 176 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 179 179 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 182 182 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 226 226 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 727 AA; 79442 MW; 9CAFL39C CRC32;

Query Match 4.18; Score 105; DB 1; Length 727;
Best Local Similarity 28.6%; Pred. No. 3.01e-01;
Matches 26; Conservative 21; Mismatches 39; Indels 5; Gaps 5;

Db 183 IRFASAIAGVDDLTGTRGN-DMOVTYIEKMFSELSGNIIIDICPVGALT-SKPYAFTA 240
QY 24 VRKISIAAGSCPLSTAGPSVYKFDQNPVGSQTFSGAHLRVFDP-STGALVDSKSYAFST 82

Db 241 RPWETRKTESIDVMDVAGSNIVVSTRT-GEV 270
QY 83 SN-DTTSAAAFVSPMNSLTNNRIVAILTSGKV 112

RESULT 12
ID NUAM_HUMAN STANDARD; PRT; 727 AA.
AC P28331;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE NADH-UBIQUINONE OXIDOREDUCTASE 75 KD SUBUNIT PRECURSOR (EC 1.6.5.3)
DE (EC 1.6.99.3) (COMPLEX I-75KD) (CI-75KD).
GN NUFS1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX CHOW W., RAGAN I., ROBINSON B.H.;
RX MEDLINE; 92037608.
RA "Determination of the cDNA sequence for the human mitochondrial
RT 75-kDa Fe-S protein of NADH-coenzyme Q reductase.";
RL EUR. J. BIOCHEM. 201:547-550(1991).
CC -!- FUNCTION: THIS IS THE LARGEST SUBUNIT OF COMPLEX I AND IT IS A
CC COMPONENT OF THE IRON-SULFUR (IP) FRAGMENT OF THE ENZYME. IT
CC MAY FORM PART OF THE ACTIVE SITE CREVICE WHERE NADH IS OXIDIZED.
CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC -!- COFACTOR: MAY BIND ONE 2FE-2S CLUSTER & ONE 4FE-4S CLUSTER.
CC -!- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 30 DIFFERENT SUBUNITS.
CC -!- SUBCELLULAR LOCATION: MATRIX AND CYTOPLASMIC SIDE OF THE
CC MITOCHONDRIAL INNER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I 75 KD SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; X61100; G38079; --
DR PIR; S16382; S16382.
DR PIR; S17854; S17854.
DR MIM; 157655; --
DR PROSITE; PS00641; COMPLEX1_75K_1; 1.
DR PROSITE; PS00642; COMPLEX1_75K_2; 1.
DR PROSITE; PS00643; COMPLEX1_75K_3; 1.
DR PFAM; PF00111; fer2; 1.
DR PFAM; PF00384; molybdopterin; 1.
KW OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION; TRANSIT PEPTIDE;
KW IRON-SULFUR; 4FE-4S.
FT TRANSIT 1 23 MITOCHONDRION (BY SIMILARITY).
FT CHAIN 24 727 NADH-UBIQUINONE REDUCTASE 75 KD SUBUNIT.
FT METAL 53 53 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 64 64 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 75 75 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 78 78 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 92 92 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 128 128 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 131 131 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 137 137 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 176 176 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 179 179 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 182 182 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 226 226 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 727 AA; 79573 MW; B9F33158 CRC32;

Query Match 4.18; Score 105; DB 1; Length 727;
Best Local Similarity 28.6%; Pred. No. 3.01e-01;
Matches 26; Conservative 21; Mismatches 39; Indels 5; Gaps 5;

Db 183 IRFASAIAGVDDLTGTRGN-DMOVTYIEKMFSELSGNIIIDICPVGALT-SKPYAFTA 240
QY 24 VRKISIAAGSCPLSTAGPSVYKFDQNPVGSQTFSGAHLRVFDP-STGALVDSKSYAFST 82

Db 241 RPWETRKTESIDVMDVAGSNIVVSTRT-GEV 270
QY 83 SN-DTTSAAAFVSPMNSLTNNRIVAILTSGKV 112

RESULT 13
ID SSP5_STRGN STANDARD; PRT; 1500 AA.
AC P16952; Q54184;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE AGGLUTININ RECEPTOR PRECURSOR.
DE SSP5 OR SSPB.
OS STREPTOCOCCUS GORDONII.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M5;
RX MEDLINE; 90236997.
RA DEMUTH D.R., GOLUB E.E., MALAMUD D.;
RT "Streptococcal-host interactions. Structural and functional analysis
RT of a Streptococcus sanguis receptor for a human salivary
RT glycoprotein.";
RL J. BIOL. CHEM. 265:7120-7126(1990).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RC STRAIN=M5;
RX MEDLINE; 96310377.
RA DEMUTH D.R., DUAN Y., BROOKS W., HOLMES A.R., MCNAB R.,
RA JENKINSON H.F.;
RT "Tandem genes encode cell-surface polypeptides SspA and SspB which
RT mediate adhesion of the oral bacterium Streptococcus gordonii to
RT human and bacterial receptors.";
RL MOL. MICROBIOL. 20:403-413(1996).
CC -!- FUNCTION: MAY BIND SALIVARY ACID RESIDUES OF SALIVARY AGGLUTININ
CC (SAG) IN A CALCIUM-DEPENDENT REACTION. THE INTERACTION OF SAG WITH
```



```
FT MOD_RES 1 1 ACETYLATION.
FT MOD_RES 254 254 METHYLATION (MONO-).
FT MOD_RES 260 260 METHYLATION (MONO-).
FT MOD_RES 372 372 METHYLATION (MONO-).
FT MOD_RES 391 391 METHYLATION (MONO-).
FT MOD_RES 392 392 METHYLATION (MONO-).
FT MOD_RES 393 393 METHYLATION (MONO-).
FT ACT_SITE 108 108 BY SIMILARITY.
SQ SEQUENCE 421 AA; 46157 MW; 2BE7B538 CRC32;

Query Match 3.9%; Score 99; DB 1; Length 421;
Best Local Similarity 21.7%; Pred. No. 1.75e+00;
Matches 15; Conservative 24; Mismatches 27; Indels 3; Gaps 3;

Db 122 EEOISRKYIOAIYKYLGSGLDIPAPDVNTDSOTMAWFL-DEYIKITGK-VDFAVETGKP 179
QY 245 DLLALGTTNISIRE-YNASNGYISSTQAEFTGQAGSWELAKEDYVVPENAVGFTIYAQT 303

Db 180 VELGGIGVR 188
QY 304 AQAGGGGMR 312
```

Search completed: Wed Sep 1 16:00:59 1999  
Job time : 24 secs.

\*\*\*\*\*  
WIREH (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:02:42 1999; MasPar time 18.01 Seconds  
Tabular output not generated. 827.821 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (1-372) from PCTUS9913024.pap (1 of 12)  
Perfect Score: 2342  
Sequence: 1 MEKFMAEFGGQVGTPTFLSE.....LPTQASSKTGKVGQEFREV 372

Scoring table: PAM 150  
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 48.649; Variance 104.828; scale 0.454

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description Pred. No.
1	1464	57.6	275 1 TLBP54 tail fiber protein gp 4.61e-250
2	128	5.0	510 2 S35028 protein P2 precursor 1.88e-03
3	124	4.9	196 2 H69647 deoxyphosphogluconate 6.17e-03
4	113	4.4	323 2 C64540 delta-aminolevulinic 1.45e-01
5	109	4.3	323 2 H71968 delta-aminolevulinic 4.36e-01
6	108	4.2	673 2 A45456 NADH dehydrogenase (u 5.72e-01
7	106	4.2	673 2 S74795 hypothetical protein 9.79e-01
8	106	4.2	2492 1 MNWVTD nonstructural polypro 9.79e-01
9	103	4.1	154 1 A40669 mitochondrial recepto 2.16e+00
10	103	4.1	277 2 B71160 probable methyl-accep 2.16e+00
11	103	4.1	300 2 S37407 glutamate dehydrogena 2.16e+00
12	103	4.1	723 2 B55926 DNA binding protein R 2.16e+00
13	105	4.1	727 2 S17854 NADH dehydrogenase (u 1.28e+00
14	105	4.1	727 2 A33552 NADH dehydrogenase (u 1.28e+00
15	105	4.1	1473 2 A35186 salivary agglutinin r 1.28e+00
16	104	4.1	2492 1 C44213 nonstructural polypro 1.66e+00
17	102	4.0	119 2 B32578 T-cell receptor beta 2.81e+00
18	102	4.0	321 2 H71729 hypothetical protein 2.81e+00
19	101	4.0	548 2 S52621 adiphosphoribosyltr 3.64e+00
20	101	4.0	1004 2 D71490 probable exodeoxyribo 3.64e+00
21	98	3.9	311 2 S66469 catechol 1,2-dioxygen 7.82e+00
22	99	3.9	421 2 S20286 glutamate dehydrogena 6.07e+00
23	99	3.9	475 1 VHVNSY nucleoprotein - Sonch 6.07e+00

24	98	3.9	572	2	S77249	hypothetical protein	7.82e+00
25	98	3.9	699	2	S62408	hypothetical protein	7.82e+00
26	100	3.9	861	2	G64087	DNA mismatch repair p	4.70e+00
27	99	3.9	945	2	S30398	aminopeptidase N homo	6.07e+00
28	99	3.9	1007	2	PN0156	glutamate receptor ch	6.07e+00
29	99	3.9	1008	2	S28858	glutamate receptor de	6.07e+00
30	100	3.9	1281	1	GNMSLL	retrovirus-related re	4.70e+00
31	99	3.9	2492	1	A44213	nonstructural polypro	6.07e+00
32	98	3.9	2567	2	A49551	filamin, Mueller cell	7.82e+00
33	99	3.9	2902	2	C71953	toxin-like outer memb	6.07e+00
34	96	3.8	163	2	D71169	hypothetical protein	1.29e+01
35	97	3.8	327	2	A60474	hypothetical protein	1.00e+01
36	97	3.8	459	2	A55484	p52(Shc) protein - mo	1.00e+01
37	97	3.8	473	2	S25776	transforming protein	1.00e+01
38	97	3.8	542	1	QREBOA	oligopeptide-binding	1.00e+01
39	96	3.8	586	2	PC6006	scaffold protein C	1.29e+01
40	96	3.8	632	2	C55926	DNA binding protein R	1.29e+01
41	96	3.8	969	2	S54594	MSU1 protein - yeast	1.29e+01
42	96	3.8	1185	2	H64526	delta-1-pyrroline-5-c	1.29e+01
43	96	3.8	1185	2	B71980	proline/pyrroline-5-c	1.29e+01
44	96	3.8	1193	2	S68218	botulinum neurotoxin	1.29e+01
45	95	3.7	394	2	A37853	flagellin, 40K - Pseu	1.65e+01

ALIGNMENTS

RESULT 1

ENTRY TLBP54 #type complete  
TITLE tail fiber protein gp35 - phase T4  
ORGANISM #formal\_name phase T4  
DATE 06-Jul-1982 #sequence\_revision 19-Apr-1996 #text\_change 29-May-1998

ACCESSIONS JT0577: A04367  
REFERENCE JT0576  
#authors Mesyanzhinov, V.V.  
#submission submitted to JIPID, April 1991  
#accession JT0577  
#status preliminary

REFERENCE A92877  
#molecule\_type protein  
#residues 1-275 #label MES  
#note Oliver, D.B.; Crowther, R.A.  
J. Mol. Biol. (1981) 153:545-568  
DNA sequence of the tail fibre genes 36 and 37 of bacteriophage T4.  
#cross-references MUID:82170495  
#accession A04367

GENETICS #molecule\_type DNA  
#residues 218-275 #label OLI  
#note the gene gp35 product was identified on the basis of its map position

GENE gp35  
#map\_position 91.6-92.3  
COMPLEX the distal half-fiber contains two molecules each of gp36 (PIR:TLBP64) and gp37 (PIR:TLBP74), and one molecule of gp35

CLASSIFICATION #superfamily phase T4 tail fiber protein gp35  
KEYWORDS structural protein; tail fiber  
SUMMARY #length 275 #molecular-weight 30237 #checksum 7602

Query Match 57.6%; Score 1464; DB 1; Length 275;  
Best Local Similarity 87.5%; Pred. No. 4.61e-250;  
Matches 244; Conservative 11; Mismatches 4; Indels 20; Gaps 1;

Db 17 FNNSLTNNRIVAILTSGKVNFPPEVVSUWLRAGT-----TSSKRA 56  
|||||  
QY 94 FNNSLTNNRIVAILTSGKVNFPPEVVSUWLRAGT-----TSSKRA 153  
|||||

Db 57 IALEHVKLSNRKSTDDYQIILWVDSLEVDGATGFPRTYVESVEFMGAVGTNNEIAR 116  
|||||

QY 154 IALEHVKLSNRKSTDDYQIILWVDSLEVDGATGFPRTYVESVEFMGAVGTNDEIAR 213  
|||||



```
Db 117 LPTSAATSKLDYLNIPGDVLYLKAQYADADLLALGTTNISIRFYNASNGYISSTQAEF 176
|||||
Qy 214 LPTSAATSKLDYLNIPGDVLYLKAQYADADLLALGTTNISIRFYNASNGYISSTQAEF 273
|||||
Db 177 TGOAGSWELKEDYVLLPQNPFPYIYHTTPQPPQGGMRNLSEFSEVRNGGISPAEFGV 236
|||||
Qy 274 TGOAGSWELKEDYVVPVENAVGFTIYQRTAQAGGMRNLSEFSEVRNGGISPAEFGV 333
|||||
Db 237 NGIRVNYICESASPPDLMVLPTQASSKTGKVGQEFREV 275
|||||
Qy 334 NGIRVNYICESASPPDLMVLPTQASSKTGKVGQEFREV 372
|||||

RESULT 2
ENTRY #type complete
TITLE protein PS2 precursor - Corynebacterium glutamicum
ORGANISM #formal_name Corynebacterium glutamicum
DATE 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change
17-Mar-1999
ACCESSIONS S35028
REFERENCE S35028
#authors Peyret, J.L.; Bavan, N.; Joliff, G.; Gulik-Krzywicki, T.;
#journal Mathieu, L.; Shechter, E.; Leblon, G.
#title Mol. Microbiol. (1993) 9:97-109
#cross-references MUID:94018618
#accession S35028
#molecule_type DNA
#residues 1-510 #label PEY
#cross-references EMBL:X69103; NID:g40488; PID:g40489
#note the authors translated the codon AAC for residue 155 as
Met
GENETICS cspB
KEYWORDS glycoprotein
FEATURE
1-30
31-510 #domain signal sequence #status predicted #label SIG\
38,58,80,187,332, #product protein PS2 #status predicted #label MAT\
353,469 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 510 #molecular-weight 55425 #checksum 3744
Query Match 5.0%; Score 128; DB 2; Length 510;
Best Local Similarity 23.3%; Pred. No. 1.88e-03;
Matches 52; Conservative 55; Mismatches 100; Indels 16; Gaps 13;

Db 134 ESNAYALKVDQET-AAFEAYRNALRDAAI-SINPDGGIN-PDTSINLLIDAANAHRD 190
|||
Qy 74 DSKSYAFSTNDTSAAFVSMNSLTNNRIVAILTSGKVNFPPEVSWLRTAGTSAPPSD 133
|||
Db 191 R--AEIE-DYAHLYTOT--DIALETPLAYAFQDLKALQAEVDADFEMLGEFGIDQ-EDG 244
|||
Qy 134 SILSRFDVSAAYTTSKRAIALEHVKLSNR-KSTDYQIILDVDFDSLEDVGATGPPRG 192
|||
Db 245 NY--VORYHLPAVEALKAEDVARVAATEPLRADSIKLNLEAKQSDVLVROLFLERATAQR 302
|||
Qy 193 TYESVEQF-MSAVGGTNDIEI-ARLPTSAAISKLS-DYNLIPGDVLYLKAQYADADLLAL 249
|||
Db 303 DTLRVVEAIFSTARYVELYENNVNENKTLRQHSALIPN 345
|||
Qy 250 GTTNISIRFYNASNGYISSTQAEFTGOAGSWELKEDY-VVYPE 291
|||

RESULT 3
ENTRY #type complete
TITLE deoxyphosphoglucuronate aldolase kgaA - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998
ACCESSIONS H69647
REFERENCE A69580
```

## #authors

Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conner, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.; Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klserr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medique, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

## #journal

Nature (1997) 390:249-256

## #title

The complete genome sequence of the Gram-positive bacterium

Bacillus subtilis.

## #cross-references

MUID:98044033

## #accession

H69647

## #status

preliminary; nucleic acid sequence not shown;

## #molecule\_type

DNA

## #residues

1-196 #label KUN

## #cross-references

GB:299115; GB:AL009126; NID:g2634478; PID:el183656;

## #experimental\_source

strain 168

## #gene

kgaA

## #CLASSIFICATION

#superfamily 2-dehydro-3-deoxyphosphoglucuronate aldolase

## #SUMMARY

#length 196 #molecular-weight 20865 #checksum 1121

## Query Match

4.9%; Score 124; DB 2; Length 196;

## Best Local Similarity

27.8%; Pred. No. 6.17e-03;

## Matches

20; Conservative 27; Mismatches 21; Indels 4; Gaps 4;

## Db

114 TPSEIMEALTFGTTKLKLPSSGVGFGPEKMLAGPQVTFITFGGTH-PSEVPDMLR-A 171

## Qy

68 STGALVDSKSYASTNDTSAAF-VSFNMLTNN-RIVAILTSGKVNFPPEVSWLR 125

## Db

172 GAGAVGVGSQGL 183

## Qy

126 GTSAPFSDSILS 137

## RESULT

4

## ENTRY

C64540

## TITLE

delta-aminolevulinic acid dehydratase - Helicobacter pylori

## ORGANISM

(strain 26695)

## DATE

09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change

## ACCESSIONS

31-Oct-1997 C64540

```

REFERENCE
#authors
A64520
Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
Richardson, D.; Dodson, R.; Khalaf, H.G.; Glodek, A.;
McKenney, K.; Fitzegerald, L.M.; Lee, N.; Adams, M.D.;
Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
Fuji, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
C.M.; Venter, J.C.
#journal
Nature (1997) 388:539-547
#title
The complete genome sequence of the gastric pathogen
Helicobacter pylori.
#cross-references
MUID:97394467
#accession
C64540
#status
preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type
DNA
#residues
1-323 #label TOM
#cross-references
GB:AE000537; GB:AE000511; NID:g2313247; PID:g2313250;
TIGR:HP0163
CLASSIFICATION
#superfamily porphobilinogen synthase
SUMMARY
#length 323 #molecular-weight 36202 #checksum 1005
Query Match 4.4%; Score 113; DB 2; Length 323;
Best Local Similarity 25.7%; Pred. No. 1.45e-01;
Matches 27; Conservative 33; Mismatches 39; Indels 6; Gaps 6;
Db 135 SVSNDKT-LEILNLOGLIAESVDILAPSNM-MDGNVLS-LRKTLDNAGYHTPTMS-Y 190
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 81 STSNDTTSAAFSFMNLTNNRIVALTSGKNFPEVSWLRTA-GTSAFSDSILSRF 139
-| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -|
Db 191 STKFASYYGPPRDVANSAPSGDRKSYMDYANQKALELSELD 235
-| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -|
QY 140 DVSAAFYSSKRAIALEHVKLSNRKSTD-DYQITLDVVDLSLED 183
-| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -|

RESULT 5
ENTRY H71968 #type complete
TITLE delta-aminolevulinic acid dehydratase - Helicobacter pylori
#formal_name Helicobacter pylori
#variety strain J99
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
05-Mar-1999
ACCESSIONS H71968
REFERENCE A71800
#authors Alm, R.A.; Ling, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.;
Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
G.F.; Trust, T.J.
#journal Nature (1999) 397:176-180
#title Genomic sequence comparison of two unrelated isolates of the
human gastric pathogen Helicobacter pylori.
#cross-references
MUID:99120557
#accession H71968
#status preliminary
#molecule_type DNA
#residues 1-323 #label ARN
#cross-references
GB:AE001453; GB:AE001439; NID:g4154651; PID:g4154663
#experimental_source strain J99
GENETICS
#gene hmbB
CLASSIFICATION
#superfamily porphobilinogen synthase
SUMMARY
#length 323 #molecular-weight 36251 #checksum 2807
Query Match 4.3%; Score 109; DB 2; Length 323;
Best Local Similarity 28.0%; Pred. No. 4.36e-01;
Matches 23; Conservative 26; Mismatches 28; Indels 5; Gaps 5;

```

---

```

Db 157 VDILAPSNM-MDGNVLS-LRKALDKAGYHTPTMS-YSTKFASYYGPPRDVANSPPSFG 213
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 104 VALTSGKNFPEVSWLRTA-GTSAFSDSILSRFDVSYAAFYSSKRAIALEHVKLS 162
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 214 DRKSYMDYANQKALELSELD 235
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 163 NRKSTD-DYQITLDVVDLSLED 183
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :

RESULT 6
ENTRY A45456 #type complete
TITLE NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain N003 -
Paracoccus denitrificans
#formal_name Paracoccus denitrificans
#journal Arch. Biochem. Biophys. (1992) 296:40-48
#title Structural features of the 66-kDa subunit of the
energy-transducing NADH-ubiquinone oxidoreductase (NDH-1)
of Paracoccus denitrificans.
#cross-references
MUID:92296779
#accession S23948
#status preliminary
#molecule_type DNA
#residues 1-673 #label XUA
#cross-references
GB:M84572; NID:g150599; PID:g150601
REFERENCE A45456
#authors Xu, X.; Matsuno-Yagi, A.; Yagi, T.
#journal Biochemistry (1993) 32:968-981
#title DNA sequencing of the seven remaining structural genes of the
gene cluster encoding the energy-transducing NADH-quinone
oxidoreductase of Paracoccus denitrificans.
#cross-references
MUID:93136200
#accession A45456
#status preliminary
#molecule_type DNA
#residues 658-673 #label XU1
#note sequence extracted from NCBI backbone (NCBIN:123409,
NCBIP:123410)
CLASSIFICATION
#superfamily NADH dehydrogenase (ubiquinone) chain 11
KEYWORDS NAD; oxidoreductase
SUMMARY
#length 673 #molecular-weight 73159 #checksum 4307
Query Match 4.2%; Score 108; DB 2; Length 673;
Best Local Similarity 26.5%; Pred. No. 5.72e-01;
Matches 22; Conservative 26; Mismatches 30; Indels 5; Gaps 5;
Db 165 VRFTTEVAGITONGTGRC-EDSEITSYLNQTLSENMQGNIDLCVPGLV-SKPYAFTA 222
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 24 VRYKTSIAGSCLSTAGSYVRFQDPNPGVSGTFSAGLHLRVFDP-STGALVDSKSYAFST 82
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 223 RPWELTKTESIDVMDALGSSIRI 245
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 83 SN-DTTSAAFYSSFMNLTNN-RI 103
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :

RESULT 7
ENTRY S74795 #type complete
TITLE hypothetical protein sll1021 - Synechocystis sp. (strain PCC
6803)
#formal_name Synechocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S74795
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;

```



```

#cross-references MUID:94119075
#accession B55926
#status preliminary
#molecule_type mRNA
#residues 1-723 #label REI
#cross-references GB:X76091; NID:g452389; PID:g452390
SUMMARY #length 723 #molecular-weight 80003 #checksum 181442390

Query Match 4.1%; Score 103; DB 2; Length 723;
Best Local Similarity 29.0%; Pred. No. 2.16e+00;
Matches 18; Conservative 17; Mismatches 25; Indels 2; Gaps 0

Db 466 VPSTLTQAINFAKSLGWL-TNMSDFPQQVIQTQKVGV-SAFATQLRRYTSLNHLAC
      :|: || |:: |:||| |:||| |:||| |:||| |:||| |:||| |:||| |:|||
Qy 103 IVALITSGKNFPEVYVWLRTAGTSAFPDSILSRFDVSAAFYATSSKKRAIALEHVVKL
      :|: || |:: |:||| |:||| |:||| |:||| |:||| |:||| |:||| |:|||

Db 524 AR 525
Qy 163 NR 164

RESULT 13
ENTRY #13
TITLE S17854 #type complete
ORGANISM NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 75K chaa
          precursor - human
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
ACCSSIONS S17854; S16382
REFERENCE S17854
          Chow, W.; Ragan, I.; Robinson, B.H.
          Eur. J. Biochem. (1991) 201:547-550
          Determination of the cDNA sequence for the human
          mitochondrial 75-kDa Fe-S protein of NADH-coenzyme
          reductase.
#cross-references MUID:92037608
#accession S17854
#molecule_type mRNA
#residues 1-727 #label CHO
#cross-references EMBL:X61100; NID:g38078; PID:g38079

GENETICS
#gene GDB:NDUFS1
#cross-references GDB:I32062; OMIM:157655
#map_position 2q33-q34
CLASSIFICATION superfamily NADH dehydrogenase (ubiquinone) chain 1
KEYWORDS iron-sulfur protein; membrane-associated complex;
          metalloprotein; mitochondrion; NAD; oxidoreductases
FEATURE
1-23 #domain transit peptide (mitochondrion) #status
      predicted #label TNP\
      #product NADH dehydrogenase (ubiquinone) 75K chaa
      #status predicted #label MAT
SUMMARY #length 727 #molecular-weight 79573 #checksum 8774

Query Match 4.1%; Score 105; DB 2; Length 727;
Best Local Similarity 28.6%; Pred. No. 1.28e+00;
Matches 26; Conservative 21; Mismatches 39; Indels 5; Gaps 0

Db 183 IRFASETAGVDDIGTTGRGN-DMQVGTYIEKMSELSGNIIDICPVGALT-SKPYPFT
      :|: || |:: |:||| |:||| |:||| |:||| |:||| |:||| |:||| |:|||
Qy 24 VRYKISIASGCSPLSTAGPSVVFQDNPNVGSQTFASGLHLRVFDP-STCALVDKSIAFYS
      :|: || |:: |:||| |:||| |:||| |:||| |:||| |:||| |:||| |:|||

Db 241 RPWETRKRTESDYMDAVGNSIVYSTRT-GEV 270
      :|: || |:: |:||| |:||| |:||| |:||| |:||| |:||| |:||| |:|||
Qy 83 SN-DTTSAAFYSPFMNSLTNNRIVALITSGKV 112

RESULT 14
ENTRY #14
TITLE A3352 #type complete
ORGANISM NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 75K chaa
          precursor - bovine
          #formal_name Bos primigenius taurus #common_name cat

```

DATE27-Feb-1990 #sequence\_revision 27-Feb-1990 #text\_change26-Feb-1999

ACCESSIONS A33552

REFERENCE A33552

#authors Runswick, M.J.; Gennis, R.B.; Fearnley, I.M.; Walker, J.E.

#journal Biochemistry (1989) 28:9452-9459

#title Mitochondrial NADH:ubiquinone reductase: complementary DNA sequence of the import precursor of the bovine 75-kDa subunit.

#cross-references MUID:90122814

#accession A33552

#molecule\_type mRNA

#residues 1-727 #label RUN

#note parts of this sequence, including the amino end of the mature protein, were determined by protein sequencing

CLASSIFICATION #superfamily NADH dehydrogenase (ubiquinone) chain II

KEYWORDS iron-sulfur protein; membrane-associated complex; metalloprotein; mitochondrion; NAD; oxidoreductase

FEATURE1-23

1-23 #domain transit peptide (mitochondrion) #status predicted #label TNP

24-727 #product NADH dehydrogenase (ubiquinone) 75K chain

#cross-references #label MAR

SUMMARY #length 727 #molecular-weight 79442 #checksum 9285

Query Match4.1%; Score 105; DB 2; Length 727;

Best Local Similarity28.6%; Pred. NO. 1.28e+00;

Matches26; Conservative21; Mismatches39; Indels5; Gaps5;

Db183 IRFASFIAGVDLGTGRGN-DMQVGTYIERFMWSELGNIIDICPVGALT-SKPYAFTA 240

Qy24 VRYKISIAAGCPLSTAGPSYVKFQDNVPVGSQTFSAGLHLRVFDP-STGALVDSKSYAFST 82

Db241 RPWETRKTESIDVMDVAGSNIVVSTRT-GEV 270

Qy83 SN-DITSAAFVFMNSLUTNNRIVAILTSGKV 112

RESULT15

ENTRY A35186 #type complete

TITLE salivary agglutinin receptor precursor - Streptococcus sanguis

ORGANISM #formal\_name Streptococcus sanguis

DATE10-Aug-1990 #sequence\_revision 10-Aug-1990 #text\_change12-Apr-1995

ACCESSIONS A35186

REFERENCE A35186

#authors Demuth, D.R.; Golub, E.E.; Malamud, D.

#journal J. Biol. Chem. (1990) 265:7120-7126

#title Streptococcal-host interactions. Structural and functional analysis of a Streptococcus sanguis receptor for a human salivary glycoprotein.

#cross-references MUID:90236997

#accession A35186

#status preliminary

#molecule\_type DNA

#residues 1-1473 #label DEM

#cross-references GB:J05418

CLASSIFICATION #superfamily surface antigen spap

SUMMARY #length 1473 #molecular-weight 162435 #checksum 9437

Query Match4.1%; Score 105; DB 2; Length 1473;

Best Local Similarity25.0%; Pred. NO. 1.28e+00;

Matches25; Conservative32; Mismatches37; Indels6; Gaps6;

Db352 NEATKQRNETAKYDAVKYKYEADLAAYKQANATNEADYQAKLAA-YQTELARVQKANA 410

Qy132 SDSILSRFDVSYAAFTYSSKR-AIALEHVKLSNRKSTDDYQITLDVWFDS-LEDVG-ATG 188

Db411 DAKATYKAVEDNKAQNAIKAEENEIKQRNAVAK-TDYE 449

Qy189 FPRGTYE-SVEQFMSAVGGTNDETARLPTSAISKLSIDYN 227

Search completed: Wed Sep 1 16:03:23 1999

Job time : 41 secs.

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on:      Wed Sep 1 16:03:42 1999;  MasPar time 15.64 Seconds
           505.902 Million cell updates/sec
Tabular output not generated.
```

```
>PCT-US99-13024-2
Title:
Description: (1-372) from PCTUS9913024.pep (1 of 12)
Perfect Score: 2542
Sequence: 1 MKKFMAEFGQYVPTPLSE.....LPTQASGKTVFGQEFREV 372
```

Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

a-geneseq35

1:part1	2:part2	3:part3	4:part4	5:part5	6:part6	7:part7
8:part8	9:part9	10:part10	11:part11	12:part12	13:part13	
14:part14	15:part15	16:part16	17:part17	18:part18		
19:part19	20:part20	21:part21	22:part22	23:part23		
24:part24	25:part25	26:part26	27:part27	28:part28		
29:part29	30:part30	31:part31	32:part32	33:part33		
34:part34	35:part35	36:part36	37:part37	38:part38		
39:part39						

Statistics: Mean 34.285; Variance 172.028; scale 0.199

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description	Pred. No.
1	300	11.8	65	19	R97371	Phage T4 ORFX gene pr	8.92e-14
2	128	5.0	510	16	R32658	P52 protein from C.me	1.50e-01
3	109	4.3	330	22	W20625	H. pylori cytoplasmic	1.00e-01
4	107	4.2	2247	25	W27126	Equine rhinovirus 1 (	1.35e+01
5	101	4.0	269	26	W38306	subtilisin 309 (BLS30	3.25e+01
6	101	4.0	269	23	W22755	subtilisin 309 (BLS30	3.25e+01
7	99	3.9	269	2	R10444	S004 mutant of the su	4.33e+01
8	99	3.9	269	2	R10445	S014 mutant of the su	4.33e+01
9	99	3.9	269	2	R10446	S020 mutant of the su	4.33e+01
10	99	3.9	270	2	R10568	S20 mutant of the sub	4.33e+01
11	99	3.9	270	2	R10441	S12 mutant of the sub	4.33e+01
12	99	3.9	270	2	R10567	S19 mutant of the sub	4.33e+01
13	99	3.9	270	2	R10442	S024 mutant of the su	4.33e+01
14	99	3.9	294	17	R88003	Delta-endotoxin MKHK	4.33e+01
15	99	3.9	571	29	W55665	H. pylori ORF hp2el09	4.33e+01
16	99	3.9	1213	30	W55735	H. pylori ORF 07ee507	4.33e+01

17	97	3.8	269	9	R46346	Subtilisin 309 Serine	5.77e+01
18	97	3.8	269	9	R46345	PB92 serine protease;	5.77e+01
19	97	3.8	473	15	R46337	Shc protein.	5.77e-01
20	97	3.8	473	21	W14005	Human Shc protein.	5.77e-01
21	97	3.8	474	17	R97243	SHC phosphotyrosine b	5.77e+01
22	97	3.8	921	25	W22863	Bacillus stearothermo	5.77e+01
23	96	3.8	1693	35	W76368	Hepatitis E virus hol	6.66e-01
24	94	3.7	269	16	R31682	Subtilisin 309 loop 5	8.84e-01
25	93	3.7	269	26	W38314	BLS309 mutant Arg170L	1.02e-02
26	93	3.7	269	23	W22766	BLS309 mutant Ser57Pr	1.02e-02
27	93	3.7	269	23	W22763	BLS309 mutant Arg170L	1.02e-02
28	93	3.7	269	26	W38317	BLS309 mutant Ser57Pr	1.02e-02
29	94	3.7	577	21	W14494	Urease subunit C.	8.84e-01
30	93	3.7	597	2	R12384	Bovine pancreatic cho	1.02e-02
31	94	3.7	664	13	R22486	Human svEGF-R11.	8.84e-01
32	94	3.7	668	28	W47037	Soluble truncated VEG	8.84e-01
33	94	3.7	764	35	W69679	KDR protein sequence.	8.84e-01
34	94	3.7	788	28	W47038	Soluble truncated VEG	8.84e-01
35	94	3.7	789	13	R62488	Truncated KLD svEGF-R	8.84e-01
36	94	3.7	848	27	W38240	VEGF receptor KDR bin	8.84e-01
37	94	3.7	1216	21	W14497	Urease protein.	8.84e-01
38	94	3.7	1356	5	R26999	Novel type III RTK en	8.84e-01
39	94	3.7	1356	32	W59275	Human KDR protein.	8.84e-01
40	91	3.6	269	26	W38310	BLS309 mutant Ser57Pr	1.35e-02
41	91	3.6	474	27	W39089	Human Shc mutant A13	1.35e-02
42	92	3.6	520	1	P46107	Neutral protease enco	1.17e+02
43	92	3.6	521	3	P35009	Sequence of neutral p	1.17e+02
44	92	3.6	543	26	W25739	Soybean protoporphyr	1.17e+02
45	91	3.6	1810	16	R94563	Chicken cvtactatin.	1.35e+02

## ALIGNMENTS

RESULT	1	
ID	R97371	standard; Protein; 65 AA.
AC	R97371;	
DE	07-JAN-1997	(first entry)
DT	Phase T4 OREX	gene product. gp34.
KW	Phase T4;	tail fibre protein; nanotechnology; nano-structure;
KK	filter;	molecular sieve.
KK	Bacteriophage T4.	
OS	WO9611947-A1.	
PN	25-APR-1996.	
PD	13-OCT-1995;	U13023.
PF	13-OCT-1994;	US-322760.
PR	(GOLD/)	GOLDBERG E. B.
PPA	Goldberg EB;	
PI	WPI; 96-221942/22.	
DR	N-PSDB: T29053.	
PT	New proteins	derived from T4 phase tail fibre proteins - that can
PT	self assemble	into nano-structure(s), useful as filters etc, also
PT	corresponding	DNA
PT	Claim 7;	Fig 7; 83pp; English.
CCC	A protein (R97371)	of unspecified function is the product of open
CCC	reading frame X	of the tail fibre protein gene region (see also
CCC	T29053) of	phase T4. This gene region also includes open reading
CCC	frames for	tail fibre proteins (see also R97370 and R97372-74).
CCC	tail fibre	proteins (native or modified) can be produced in large
CCC	quantities	in microbial cells and used as bulding blocks of strong
CCC	stable	nanosstructures.
CC	Sequence	65 AA;
CC	Sequence	65 AA;

RESULT 2  
ID R32658 standard; Protein; 510 AA.

```

AC R32658;
DE 18-JUN-1993 (first entry)
DE PS2 protein from C.melasscola ATCC 17965.
KW NADPH dependent glutamate dehydrogenase; Corynebacterium glutamicum;
KW cs2 gene; secretory peptide.
OS Corynebacterium melasscola ATCC 17965.
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= signal
FT /note= "for protein secretion"
FT cleavage_site 26..30
FT /note= "putative"
FT protein 31..510
FT /label= PS2
FT
FT W09303158-A.
PD 18-FEB-1993.
PF 29-JUL-1992; F00744.
PR 30-JUL-1991; FR-009652.
PR 02-AUG-1991; FR-009870.
PA (ORSA-) ORSAN.
PI Duchiron F, Guyonvarch A, Joliff G, Leblon G, Relano P;
PI Renaud M;
DR WPI; 93-076515/09.
DR N-PSDB: Q37239.
PT Protein expression and secretion system in corynebacteria -
PT transformed with cassette contg. signal sequence of protein PS1
PT or 2 in chromosome or plasmid DNA region, for repetitive
PT synthetic aminoacid polymer and polypeptide prodn.
PS Claim 1; Fig 12; 9pp; French.
CC N-terminal signal peptide of PS2 is useful in recombinant protein
CC production. It results in secretion of the protein from the
CC transformant. Since PS2 is associated with the bacterial cell wall,
CC a portion of the expressed product will remain attached to the wall
CC and can be recovered (without lysis) by treatment with a surfactant.
SO Sequence 510 AA;

Query Match 5.0%; Score 128; DB 6; Length 510;
Best Local Similarity 23.3%; Pred. No. 5.54e-01;
Matches 52; Conservative 55; Mismatches 100; Indels 16; Gaps 13;

Db 134 esvayalkvdeat-aafeyarnalrdaai-sinpdgsin-pdtsinllidaanaanrtid 190
Qy 74 DSKSYAFSTSDTSAFVSEFMNSLTNNRIVAILTSGRVNFPPEVWSLRTAGTSAPPSD 133
Db 191 r-aeie-dvahltyqt--dialetpqlayafqdlkalqaevdadfewlgefgidq-edg 244
Qy 134 SILSKFVSYAAFTISSRRALAEHVKLSNR-KSTDDYQTILDVVFDSDLEDVGGTGFPRG 192
Db 245 ny--vqryhlpavealkaevdarvaaielpiradsiaakleaqksdvlvrqlflerataqr 302
Qy 193 TYESVEQF-MSAVGGTNDI-ARLPTSAAISKLS-DYNLIPGDVLYLKAQLYADADLLAL 249
Db 303 dclrvaeifstaryvelyenvenvntklrhysaallpn 345
Qy 250 GTTNISIRFYNASNGYISSTQAEFTQGAGSWELKEDY-VVYVE 291

RESULT 3
ID W20625 standard; Protein: 330 AA.
AC W20625;
DE 14-JUL-1997 (first entry)
DE H. pylori cytoplasmic protein, O2ae11612orf26.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacterium; life cycle; activator;
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW diagnosis.
OS Helicobacter pylori.
PN W09640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR ) ASTRA AB.

```

```

PI Berglindh OT, Smith D, Mellgaard BL;
DR WPI; 97-052306/05.
DR N-PSDB: T67878.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 61; Page 1051-52; 1481pp; English.
CC This sequence represents a H. pylori cytoplasmic protein involved in
CC cofactor metabolism.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori life cycle activators or inhibitors.
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 330 AA;

Query Match 4.3%; Score 109; DB 22; Length 330;
Best Local Similarity 28.0%; Pred. No. 1.00e+01;
Matches 23; Conservative 26; Mismatches 28; Indels 5; Gaps 5;

Db 164 vdilaplanm-mdgnvls-lrkalkdkgfyhtplms-ystkfassygpfdrdvanspsfg 220
Qy 104 VAILTSGRVNFPPEVWSLRTA-CTSAPPSDILSRFDVSYAAFTSSKRAIALEHVKLS 162
Db 221 drksymdyanqkeallesled 242
Qy 163 NRKSTD-DYQITLDVWFDSLED 183

RESULT 4
ID W27126 standard; Protein: 2247 AA.
AC W27126;
DE 11-FEB-1998 (first entry)
DE Equine rhinovirus 1 (ERHV1), polyprotein amino acid sequence.
DE Equine rhinovirus 1; ERHV1; foot-and-mouth disease virus; vaccine; horse;
DE diagnosis; antigens; polyprotein; enzyme-linked immunosorbent assay;
KW recombinant protein.
OS Equine rhinovirus 1.
FH Key Location/Qualifiers
FT Cleavage_site 207..208
FT Cleavage_site 287..288
FT Cleavage_site 517..518
FT Cleavage_site 743..744
FT Cleavage_site 991..992
FT Cleavage_site 1007..1008
FT Cleavage_site 1143..1144
FT Cleavage_site 1158..1159
FT Cleavage_site 1552..1553
FT Cleavage_site 1577..1578
FT Cleavage_site 1782..1783
FT Peptide 1..207
FT /label= L
FT Peptide 208..287
FT /label= VP4
FT Peptide 288..517
FT /label= VP2
FT Peptide 518..743
FT /label= VP3
FT Peptide 744..991
FT /label= VP1
FT Peptide 992..1007
FT /label= 2A
FT Peptide 1008..1143
FT /label= 2B
FT Peptide 1144..1458

```

FT Peptide /label= 2C  
 FT 1459..1552  
 FT /label= 3A  
 FT 1553..1577  
 FT /label= 3B  
 FT 1578..1782  
 FT /label= 3C  
 FT 1783..2246  
 FT /label= 3D  
 PN W09722701-A1.  
 PD 26-JUN-1997.  
 PF 18-DEC-1996: AU0815.  
 PR 18-DEC-1995: AU-007201.  
 PA (UYME ) UNIV MELBOURNE.  
 PI Crabb BS, Feng L, Studdert MJ;  
 DR WPI: 97-341692/31.  
 DR N-PSDB: T83178.  
 DR Genomic sequence of equine rhinovirus 1 - and derived proteins or  
 PT virus-like particles, useful in vaccines and as diagnostic agents  
 PS Claim 2: Pages 30-32; 60pp; English.  
 CC The present sequence represents the amino acid sequence of the  
 CC polypeptide of equine rhinovirus 1 (ERHV1). The taxonomic status  
 CC of ERHV1 is unclear, as physicochemical studies have shown that the  
 CC nucleic acid density and base composition of ERHV1 differs from other  
 CC rhinoviruses. To this end, the nucleotide sequence encoding the  
 CC polypeptide of ERHV1 was deduced. Analysis of this sequence suggests that  
 CC ERHV1 is more closely related to foot-and-mouth disease virus. Individual  
 CC ERHV1 proteins can be used to make vaccines to protect horses (and  
 CC possibly other animals) against ERHV1. Oligonucleotide primers and probes  
 CC can be used for diagnosis of ERHV1 or related viruses, while antigens of  
 CC the ERHV1 polypeptide can be used to detect ERHV1-specific antibodies in  
 CC the blood, particularly in enzyme-linked immunosorbent assay. They can  
 CC differentiate between infected animals and those vaccinated with ERHV1  
 CC vaccines (the infected animals will have antibodies reactive with  
 CC non-capsid proteins but vaccinated animals will not). Fragments of the  
 CC DNA sequence represent individual genes of the virus and can be  
 CC expressed in host systems to produce recombinant proteins. Virus like  
 CC particles containing the individual ERHV1 proteins, can also be used as  
 CC vectors for delivering therapeutic or other useful agents, including  
 CC vaccinating epitopes from other pathogens or reproductive hormones.  
 SQ Sequence 2247 AA;  
 Query Match 4.2%; Score 107; DB 25; Length 2247;  
 Best Local Similarity 23.4%; Pred. No. 1.35e+01;  
 Matches 22; Conservative 33; Mismatches 32; Indels 7; Gaps 7;  
 Db 1041 tpspfykamkflklatlavaamrtkdppvvmliafd-gle-vfdtgffsfyfqklqp 1098  
 QY 142 SYAAFTSSKRAIALEHVKLSNRKSTDDYQTILDVY-FDSLEDVGATGPRGTY-ESVEQ 199  
 Db 1099 ymktpgkisdvltdaataaqlpk-gvysfvss 1131  
 QY 200 FMSAVGNTNDE-IARLPTSNA-ISKLSIDNLIIPG 231  
 RESULT 5  
 ID W38306 standard; protein; 269 AA.  
 AC W38306;  
 DT 27-MAR-1998 (first entry)  
 DE subtilisin 309 (BLS309) mutant Arg170Phe (in BASBPN numbering).  
 KW Variant; mutant; subtilase; enzyme; BASBPN; storage stability;  
 KW wash performance; detergent; hydrophobic domain; subtilisin 309;  
 KW BLS309.  
 OS Bacillus lentus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Misc\_difference 164  
 FT /note= "wild type Arg replaced by Phe"  
 PN W09634935-A2.  
 PD 07-NOV-1996.  
 PF 12-APR-1996: E01610.  
 PR 05-MAY-1995: EP-201161.  
 PA (UNIL ) UNILEVER NV.

PA (UNIL ) UNILEVER PLC.  
 PI Bauditz P, Markvardsen P, Sierkstra LN, Von Der Osten C,  
 DR Klugkist J;  
 DR WPI: 96-506142/50.  
 PT Detergent compositions comprising Subtilisin variants - which  
 PT exhibit improved stability and/or wash performance  
 PS Claim 1: Page -: 91pp; English.  
 CC The present subtilase BLS309 variant, which has improved storage  
 CC stability and/or improved performance in detergents, has a residue  
 CC in the vicinity a wild type hydrophobic domain substituted with a  
 CC more hydrophobic residue. The hydrophobic domain comprises the  
 CC residues Pro129, Pro131, Ile165, Tyr167, Tyr171 of BLS309 (in BASBPN  
 CC numbering). Detergent compositions containing the mutant enzyme  
 CC have improved storage stability, while retaining or improving their  
 CC wash performance. This variant has a wash performance improvement  
 CC factor of 3.3 compared to wild type BLS309.  
 CC N.B. Sequence not given in the specification, but constructed using  
 CC the wild type BLS309 sequence, Genbank A08332.  
 SQ Sequence 269 AA;  
 Query Match 4.0%; Score 101; DB 26; Length 269;  
 Best Local Similarity 32.8%; Pred. No. 3.25e+01;  
 Matches 19; Conservative 21; Mismatches 14; Indels 4; Gaps 4;  
 Db 159 isypafyanamavgatdq-n-nnrasfsgyagldivapgv-nvqst-ypgstvasln 212  
 QY 141 VSYAAFTSSKRAIALEHVKLSNRKSTDDYQTILDVY-FDSLEDVGATGPRGTYESVE 198  
 RESULT 6  
 ID W22755 standard; protein; 269 AA.  
 AC W22755;  
 DT 11-SEP-1997 (first entry)  
 DE subtilisin 309 (BLS309) mutant Arg170Phe (in BASBPN numbering).  
 KW Variant; mutant; subtilase; enzyme; BASBPN; storage stability;  
 KW wash performance; detergent; hydrophobic domain; subtilisin 309;  
 KW BLS309.  
 OS Bacillus lentus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Misc\_difference 164  
 FT /note= "wild type Arg replaced by Phe"  
 PN W09634946-A1.  
 PD 07-NOV-1996.  
 PF 02-MAY-1996: DK0207.  
 PR 05-MAY-1995: DK-000519.  
 PR 12-APR-1996: DK-000421.  
 PA (NOVO ) NOVO-NORDISK AS.  
 PI Bauditz P, Klugkist J, Markvardsen P, Sierkstra LN;  
 PI Von Der Osten C;  
 DR WPI: 96-506152/50.  
 PT A subtilase variant with amino acid substitutions in or near  
 PT hydrophobic domain - has improved storage stability and/or improved  
 PT performance in detergents, whilst retaining or improving its wash  
 PT performance  
 PS Claim 1: Page -: 91pp; English.  
 CC The present subtilase BLS309 variant, which has improved storage  
 CC stability and/or improved performance in detergents, has a residue  
 CC in the vicinity a wild type hydrophobic domain substituted with a  
 CC more hydrophobic residue. The hydrophobic domain comprises the  
 CC residues Pro129, Pro131, Ile165, Tyr167, Tyr171 of BLS309 (in BASBPN  
 CC numbering). Detergent compositions containing the mutant enzyme  
 CC have improved storage stability, while retaining or improving their  
 CC wash performance. This variant has a wash performance improvement  
 CC factor of 3.3 compared to wild type BLS309.  
 CC N.B. Sequence not given in the specification, but constructed using  
 CC the wild type BLS309 sequence, Genbank A08332.  
 SQ Sequence 269 AA;  
 Query Match 4.0%; Score 101; DB 23; Length 269;  
 Best Local Similarity 32.8%; Pred. No. 3.25e+01;  
 Matches 19; Conservative 21; Mismatches 14; Indels 4; Gaps 4;





CC detergent liquids, or detergent bars. See also R10440-41, R10499-  
CC 500, R10567-68, EP-405901 and W09100345.  
SQ Sequence 270 AA;

Query Match 3.9%; Score 99; DB 2; Length 270;  
Best Local Similarity 31.0%; Pred. No. 4.33e+01;  
Matches 18; Conservative 22; Mismatches 14; Indels 4; Gaps 4;

Db 160 isypayyanamavgatdq-n-nnrassfsqygaeldivapgv-nvqst-ypgstysasin 213  
QY 141 VSYAAFTSSKRAIALEHVKLSNRKSTDDYQITLDVVFDSLEDVGATGPRGTYESVE 198

RESULT 11  
ID R10441 standard; Protein; 270 AA.  
AC R10441;  
DT 19-MAR-1991 (first entry)  
DE S12 mutant of the subtilisin 309 enzyme.  
KW Subtilisin mutant; S12; detergent.  
OS Subtilisin 309.  
PN EP-405902-A.  
PD 02-JAN-1991.  
PF 26-JUN-1990; 306953.  
PR 26-JUN-1989; GB-014605.  
PR 07-JUL-1989; GB-015660.  
PR 26-JUN-1990; EP-306953.  
PA (UNIL ) UNILEVER PLC.  
PI Casteleijn E, Egmond MR, Haverkamp J, Marugg JD, Mooren AT;  
DR WPI; 91-009218/02.  
PT Detergent compns. contg. enzyme - comprising mutated subtilisin  
PT protease  
PS Disclosure; Page 6; 10pp; English.  
CC S12) G195E + R170Y + K251E.  
CC The wild-type sequence is obtained from the International Patent  
CC Application NO. PCT/DK 88/00002. The mutant protease is used in  
CC certain detergent compositions formulated as detergent powders contg.  
CC phosphate or zeolite builder, aqueous detergent liquids, nonaqueous  
CC detergent liquids, or detergent bars. See also R10440-41, R10499-  
CC 500, R10567-68, EP-405901 and W09100345.  
SQ Sequence 270 AA;

Query Match 3.9%; Score 99; DB 2; Length 270;  
Best Local Similarity 31.0%; Pred. No. 4.33e+01;  
Matches 18; Conservative 22; Mismatches 14; Indels 4; Gaps 4;

Db 160 isypayyanamavgatdq-n-nnrassfsqygaeldivapgv-nvqst-ypgstysasin 213  
QY 141 VSYAAFTSSKRAIALEHVKLSNRKSTDDYQITLDVVFDSLEDVGATGPRGTYESVE 198

RESULT 12  
ID R10567 standard; Protein; 270 AA.  
AC R10567;  
DT 19-MAR-1991 (first entry)  
DE S19 mutant of the subtilisin 309 enzyme.  
KW Subtilisin mutant; S19; detergent.  
OS Subtilisin 309.  
PN EP-405902-A.  
PD 02-JAN-1991.  
PF 26-JUN-1990; 306953.  
PR 26-JUN-1989; GB-014605.  
PR 07-JUL-1989; GB-015660.  
PR 26-JUN-1990; EP-306953.  
PA (UNIL ) UNILEVER PLC.  
PI Casteleijn E, Egmond MR, Haverkamp J, Marugg JD, Mooren AT;  
DR WPI; 91-009218/02.  
PT Detergent compns. contg. enzyme - comprising mutated subtilisin  
PT protease  
PS Disclosure; Page 6; 10pp; English.  
CC S19) G195E + R170Y + H120D + K235L.  
CC The wild-type sequence is obtained from the International Patent  
CC Application NO. PCT/DK 88/00002. The mutant protease is used in  
CC certain detergent compositions formulated as detergent powders contg.

CC phosphate or zeolite builder, aqueous detergent liquids, nonaqueous  
CC detergent liquids, or detergent bars. See also R10440-41, R10499-  
CC 500, R10567-68, EP-405901 and W09100345.  
SQ Sequence 270 AA;

Query Match 3.9%; Score 99; DB 2; Length 270;  
Best Local Similarity 31.0%; Pred. No. 4.33e+01;  
Matches 18; Conservative 22; Mismatches 14; Indels 4; Gaps 4;

Db 160 isypayyanamavgatdq-n-nnrassfsqygaeldivapgv-nvqst-ypgstysasin 213  
QY 141 VSYAAFTSSKRAIALEHVKLSNRKSTDDYQITLDVVFDSLEDVGATGPRGTYESVE 198

RESULT 13  
ID R10442 standard; Protein; 270 AA.  
AC R10442;  
DT 18-MAR-1991 (first entry)  
DE S024 mutant of the subtilisin 309 enzyme.  
KW Subtilisin 309 mutant; detergent; S024.  
OS Bacillus subtilis 309.  
PN EP-405901-A.  
PD 02-JAN-1991.  
PF 26-JUN-1990; 306952.  
PR 26-JUN-1989; GB-014604.  
PR 07-JUL-1989; GB-015660.  
PR 26-JUN-1990; EP-306952.  
PA (UNIL ) UNILEVER PLC.  
PI Casteleijn E, Egmond MR, Haverkamp J, Marugg JD, Mooren AT;  
DR WPI; 91-009217/02.  
PT Enzymatic detergent compns. - contg. a mutant subtilisin  
PT protease with an isoelectric pH lower or higher than the parent  
PT protease  
PS Claim 9; Page 47; 57pp; English.  
CC S024) \*36D + H120D + R170Y + G195E + K235L + K251E.  
CC The mutant enzyme is produced by mutating the gene of subtilisin 309  
CC protease and expressing the mutated gene in suitable hosts. The  
CC enzyme exhibits improved wash performance in comparison with their  
CC wild type parent enzyme. The enzyme is well suited for use in  
CC detergent compositions. The wild type sequence is taken from the  
CC International Patent Application NO. PCT/DK 88/00002.  
CC See also R10442-56, EP-405902 and W09100345.  
SQ Sequence 270 AA;

Query Match 3.9%; Score 99; DB 2; Length 270;  
Best Local Similarity 31.0%; Pred. No. 4.33e+01;  
Matches 18; Conservative 22; Mismatches 14; Indels 4; Gaps 4;

Db 160 isypayyanamavgatdq-n-nnrassfsqygaeldivapgv-nvqst-ypgstysasin 213  
QY 141 VSYAAFTSSKRAIALEHVKLSNRKSTDDYQITLDVVFDSLEDVGATGPRGTYESVE 198

RESULT 14  
ID R88003 standard; Protein; 294 AA.  
AC R88003;  
DT 02-APR-1996 (first entry)  
DE Delta-endotoxin MKHK.  
KW Lepidoptera; insecticide; Bacillus thuringiensis; B.t.; MKHKH;  
KW lepidoptera; coleoptera; pest control; biopesticide.  
OS Bacillus thuringiensis.  
PN W09535378-A1.  
PD 28-DEC-1995.  
PF 15-JUN-1995; U07537.  
PR 22-JUN-1994; US-264100.  
PR 10-NOV-1994; US-337358.  
PA (NOVO ) NOVO-NORDISK ENTOTECH INC.  
PI Adams L, Liu C, Lufburrow PA, Thomas MD;  
DR WPI; 96-058415/06.  
PT New delta-endotoxins from Bacillus thuringiensis strains - used in  
PT insecticidal compns. against lepidopteran and coleopteran pests  
PS Claim 2; Page 33-34; 53pp; English.  
CC R88002 and R88003 represent delta-endotoxins obtained from Bacillus

thuringiensis (B.t.). This sequence represents the MKHHK delta-endotoxin, and has a molecular weight of 33 kDa. Delta-endotoxins have molecular weights ranging from 27-140 kDa, and kill insect larvae upon ingestion. These sequences can be used in insecticidal compositions, which can be used to control lepidopteran and coleopterian pests. These compositions can optionally contain spores from a biologically pure B.t. strain. The compositions also contain at least two other delta-endotoxins, each having a molecular weight of 130 kDa. The delta-endotoxins increase the spectrum of available biopesticides, particularly to any given insect pest.

SQ Sequence 294 AA;

Query Match 3.9%; Score 99; DB 17; Length 294;  
Best Local Similarity 26.0%; Pred. No. 4.33e+01;  
Matches 25; Conservative 28; Mismatches 38; Indels 5; Gaps 5;  
Db 119 tqsltvskfsfnfpvagaennisfsvgeqnlstekttestlmripqpvsvrprt 178  
Qy 69 TCAAL-VDSK-SYAFSTSDNTTSAAF-VSFMSLTNNRIVAILTSGKVNFPPEVWS-WLRT 124  
Db 179 akrveislfeilaiprigneisgfvgtlptltsnsh 214  
Qy 125 AGTSAFP-SDSILSRFDYSYAAFTYSSKRALEHV 159

RESULT 15  
ID W55665 standard; Protein: 571 AA.

AC W55665;  
DT 24-JUN-1998 (first entry)  
DE H. pylori ORF hpzel0911\_4882027\_c2\_87 secreted protein.  
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope; identification; binding compound; bacteria; life cycle; activator; inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.  
OS Helicobacter pylori.  
PN W09737044-A1.  
PD 09-OCT-1997.  
PF 27-MAR-1997; U05223.  
PR 06-DEC-1996; US-761318.  
PR 29-MAR-1996; US-625811.  
PR 02-APR-1996; US-758731.  
PR 25-OCT-1996; US-736905.  
PR 28-OCT-1996; US-738859.  
PA (ASTR) ASTRA AB.  
PI Alm RA, Smith D;  
DR WPI; 97-503122/46.  
DR N-PSDB: V25074.  
DT Helicobacter pylori nucleic acid sequences and encoded polypeptide(s) - useful in vaccines to treat or prevent H. pylori infection and for diagnosis of H. pylori infection  
PS Disclosure: Pages 871-872; 1145pp; English.  
CC This is the sequence of a Helicobacter pylori secreted protein.  
CC The protein may be used in a vaccine to prevent or treat H. pylori infection or to identify H. pylori polypeptide binding compounds, useful as potential H. pylori life cycle activators or inhibitors. The DNA and probes derived from it may be used for the identification of H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic acid sequences complementary to the DNA act as antisense sequences and can be used to prevent the translation of H. pylori mRNA. Antibodies against the protein can be used in immunoassays to evaluate the abundance and distribution of H. pylori-specific antigens. The genomic sequence of H. pylori (ATCC 55679) was determined from overlapping contigs generated by mechanically shearing the bacterial DNA. The sequences were analysed for ORF of at least 180 nucleotides, and the predicted coding regions defined by computer evaluation. To identify likely H. pylori antigens for vaccine development, the amino acid sequences predicted from various ORF were analysed for significant homology to other known or exported membrane proteins. Having identified and determined the sequences of interest, particular regions can be isolated from H. pylori by PCR amplification for recombinant polypeptide production, e.g. in E. coli hosts.

SQ Sequence 571 AA;

Query Match 3.9%; Score 99; DB 29; Length 571;

Best Local Similarity 21.6%; Pred. No. 4.33e+01;  
Matches 30; Conservative 43; Mismatches 58; Indels 8; Gaps 7;  
Db 190 lnfnnnqfngsfeigakntifn-nan-fnnstsfnnssatts--fvgdftnansnl 245  
Qy 44 VKF-QDNVPVGSQTFSAGLHLRFDPSTGALVDSKSYAFSTNDTTSAAAFVS-FMNSLTNN 101  
Db 246 qiagnavfngstngsqntanfnntgsvniagnatfdnv-vfnsptntsvkgkvtlnnltl 304  
Qy 102 RIVAILTSGKVNFPPEVSWLRTAGTSAPPSDSILSRFDVSYAAFTYSSKRAIALEHVKL 161  
Db 305 knlnaplsfgdg-tivfsa 322  
Qy 162 SNRKSTDDYQITLDVWFDS 180

Search completed: Wed Sep 1 16:04:42 1999  
Job time : 60 secs.

\*\*\*\*\*  
M P S R C H (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:01:19 1999; MasPar time 25.10 Seconds  
Tabular output not generated. 808.836 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (1-372) from PCTUS9913024.pep (1 of 12)  
Perfect Score: 2342  
Sequence: 1 MEKFAEFGGQVOTPFLE.....LPTQASSKTGKVGQEFREV 372

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sprenb19  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle  
9:sp-phase 10:sp-plant 11:sp-rodent 12:sp\_unclassified  
13:sp-vertebrate 14:sp\_virus

Statistics: Mean 47.705; Variance 93.090; scale 0.512

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	397	15.6	58	9	Q38400 TAIL FIBER GENE 35 (FR	1.12e-52
2	391	15.4	58	9	Q38425 TAIL FIBER GENE 35 (FR	1.70e-51
3	382	15.0	58	9	Q38393 TAIL FIBER GENE 35 (FR	9.94e-50
4	128	5.0	510	2	Q04985 P52 PROTEIN PRECURSOR.	3.35e-04
5	111	4.1	373	5	Q61942 T24A6.11 PROTEIN.	7.82e-02
6	113	4.4	227	14	Q66775 POLYPROTEIN.	4.22e-02
7	113	4.4	2248	14	Q66774 POLYPROTEIN.	4.22e-02
8	113	4.4	2497	14	Q88578 NONSTRUCTURAL POLYPROT	4.22e-02
9	109	4.3	1007	4	Q43424 GLUTAMATE RECEPTOR DEL	1.44e-01
10	107	4.2	920	2	Q45664 SBSB GENE.	2.63e-01
11	106	4.2	1300	11	Q08906 REVERSE TRANSCRIPTASE	3.59e-01
12	107	4.2	2247	14	Q39818 POLYPROTEIN.	2.63e-01
13	106	4.2	2493	14	Q09163 NONSTRUCTURAL POLYPROT	3.55e-01
14	106	4.2	2493	14	Q66594 NONSTRUCTURAL POLYPROT	3.55e-01
15	106	4.2	2493	14	Q66592 PUTATIVE NONSTRUCTURAL	3.55e-01
16	103	4.1	277	1	Q58196 277AA LONG HYPOTHETICA	8.59e-01
17	103	4.1	325	2	Q88073 PUTATIVE TRANSCRIPTION	8.59e-01
18	105	4.1	1546	2	Q45996 SCAFFOLDING PROTEIN PR	4.78e-01
19	101	4.0	1548	10	Q39000 AMIDOPHOSPHORIBOSYLTRA	1.53e+00
20	101	4.0	997	5	Q44825 H17B01.4 PROTEIN.	1.53e+00

21 101 4.0 1004 2 084646 EXODEOXYRIBONUCLEASE V 1.53e+00  
22 101 4.0 1577 2 054178 GLUCOSYLTRANSFERASE. 1.53e+00  
23 98 3.9 203 10 049136 CELL DIVISION PROTEIN 3.60e+00  
24 98 3.9 311 2 052041 CATECHOL 1,2-DIOXYGENA 3.60e+00  
25 98 3.9 321 5 077113 PERIOD (FRAGMENT). 3.60e+00  
26 98 3.9 383 5 016306 T21H3.5 PROTEIN. 3.60e+00  
27 100 3.9 466 11 061787 ORF 2. 2.04e+00  
28 98 3.9 572 2 073543 HYPOTHETICAL 62.9 KD P 3.60e+00  
29 100 3.9 823 10 039594 FUS1 PROTEIN. 2.04e+00  
30 100 3.9 863 4 014689 KIAA0184 PROTEIN (FRAG 2.04e+00  
31 99 3.9 1007 11 061825 GLUTAMATE RECEPTOR CHA 2.72e+00  
32 99 3.9 1007 11 062841 GLUTAMATE RECEPTOR DEL 2.72e+00  
33 99 3.9 1008 11 063226 GLUTAMATE RECEPTOR SUB 2.72e+00  
34 100 3.9 1281 11 088915 L1 RETROTRANSPOSON, TF 2.04e+00  
35 99 3.9 1281 11 088914 L1 RETROTRANSPOSON, TF 2.72e+00  
36 99 3.9 1281 11 054850 ENDONUCLEASE/REVERSE T 2.72e+00  
37 99 3.9 1281 11 088913 L1 RETROTRANSPOSON, TF 2.72e+00  
38 100 3.9 1300 11 061785 LIMD-A13 REPETITIVE SE 2.04e+00  
39 100 3.9 1693 14 081876 COMPLETE GENOME SEQUEN 2.04e+00  
40 98 3.9 2567 13 090574 FILAMIN. 3.60e+00  
41 97 3.8 311 2 051960 CATECHOL OXYGENASE. 4.76e+00  
42 97 3.8 413 2 048595 PLASMID\_PTR2030 LLAI M 4.76e+00  
43 97 3.8 476 4 075674 TOM1-LIKE PROTEIN. 4.76e+00  
44 97 3.8 583 4 015290 P66SHC. 4.76e+00  
45 97 3.8 1350 5 015884 INVARIANT SURFACE GLYC 4.76e+00

ALIGNMENTS

RESULT 1  
ID Q38400 PRELIMINARY; PRT; 58 AA.  
AC Q38400;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE TAIL FIBER GENE 35 (FRAGMENT).  
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE;  
OS BACTERIOPHAGE OX2.  
OC T4-LIKE PHAGES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 85215500.  
RA RIEDE I., DREXLER K., ESCHBACH M.L.;  
RT "The nucleotide sequences of the tail fiber gene 36 of bacteriophage  
RT T2 and of genes 36 of the T-even type Escherichia coli phages K3 and  
RT OX2".  
RL NUCLEIC ACIDS RES. 13:605-616(1985).  
DR EMBL; X01753; E7867; -.  
FT NON\_TER 1  
SQ SEQUENCE 58 AA; 6206 MW; 29127C56 CRC32;  
Query Match 15.6%; Score 397; DB 9; Length 58;  
Best Local Similarity 98.3%; Pred. No. 1.12e-52;  
Matches 57; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
DB 1 SFSEVSRNGGSKPAEFGVNGIRVNYVCESASPPDMLVLPQTQASSKTGKVGQEFREV 58  
QY 315 SFSEVSRNGGSKPAEFGVNGIRVNYVCESASPPDMLVLPQTQASSKTGKVGQEFREV 372  
RESULT 2  
ID Q38425 PRELIMINARY; PRT; 58 AA.  
AC Q38425;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE TAIL FIBER GENE 35 (FRAGMENT).  
OC BACTERIOPHAGE T2.  
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE;  
OC T4-LIKE PHAGES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 85215500.

	Query Match	4.4%	Score 111;	DB 5;	Length 373;
	Best Local Similarity	24.0%	Pred. NO. 7.82e-02;		
	Matches	36;	Conservative	44;	Mismatches 60; Indels 10; Gaps 9;
Db	157	SPEV-TRVSTSTP-VFOSSAYDMSDADRRFFMTYITIEFLNLDPMKLEHSDKQWLLON	214		
QY	115	PPEWVKWLRTAGTSAPPSDSI-USRDEVYAAFTTSKRAIALEHY-KL-SNRKST-DD	169		
Db	215	FSKATLLFNLSLTRVGNNDRLTTPGGHIEIVPDLMTFLFNVSLSFTQIRSVRVSKLIBL	274		

```
QY   170 YQTILDVVFSLDV-CATG-FP-RGTYESVEGFMSAVGTNDENDEIARLPTSAAISKLSDY 222
Db   275 EITDEFLLVTILFCDDPAIGSSHAISI 304
      :: : | : | : |:: |||
QY   227 NLIPGDVLKAOLYADADALLAGTTNISI 256


RESULT    6
ID Q66775 PRELIMINARY; PRT; 2227 AA.
AC Q66775;
DT DT 01-NOV-1996 (TREMBREL. 01, CREATED)
DT DT 01-NOV-1996 (TREMBREL. 01, LAST SEQUENCE UPDATE)
DE DE 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DS POLYPROTEIN.
OS EQUINE RHINOVIRUS TYPE 1.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
CC APHTHOVIRUS.
[1]
RN RN SEQUENCE FROM N.A.
RC STRAIN-PERV;
RX MEDLINE; 96332508.
RA WUTZ G., AUER H., NOWOTNY N., SKERN T., KUECHLER E.;
RT "Equine rhinovirus serotypes 1 and 2: relationship to each other and
RL rt to aphthoviruses and cardioviruses." ;
RJR J. GEN. VIROL. 77:1719-1730(1996).
DR EMBL; X96870; E236020; - .
DR PFAM; PF00073; rlv; 2.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
KW POLYPROTEIN.
FT FT CHAIN          189     268       VP4.
FT FT CHAIN         269     498       VP2.
FT FT CHAIN        499     724       VP3.
FT FT CHAIN        725     972       VPI.
FT FT CHAIN        973     988       2A..
FT FT CHAIN        989    1124       2B..
FT FT CHAIN       1125    1439       2C..
FT FT CHAIN       1440    1534       3A..
FT FT CHAIN       1535    1558       3B (VPG)..
FT FT CHAIN       1559    1763       3C (PROTEINASE)..
FT FT CHAIN       1764    2227       3D (RNA POLYMERASE)..
SQ SEQUENCE      2227 AA; 247166 MW; A4BIC493 CRC32;

Query Match              4.4%; Score 113; DB 14; Length 2227;
Best Local Similarity 24.5%; Pred.No. 4.22e-02;
Matches 23; Conservative 33; Mismatches 31; Indels 7; Gaps

Db   1022 TPSPFYKTAKMLFKLATLVATTMRTKDPVVVMLIADF-GLE-VFDTGFFSFYSFOEKLP 1079
      :||:|||:|:|:|||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~
QY   142 SYAFYTSSRRALALEHVLNSRKSTDDTQTLDIV-FDSLVDVGATGPFRGTY-ESVEQ 199
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~

Db   1080 YMKTIPKISELTDAATAAQAIKP-GVTSFVS 1112
      |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~
QY   200 FMSAVGTNDEN-IARLPTSAA-ISKLSDYNLI PG 231
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~


RESULT    7
ID Q66774 PRELIMINARY; PRT; 2248 AA.
AC Q66774;
DT DT 01-NOV-1996 (TREMBREL. 01, CREATED)
DT DT 01-NOV-1996 (TREMBREL. 01, LAST SEQUENCE UPDATE)
DE DE 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DS POLYPROTEIN.
OS EQUINE RHINOVIRUS TYPE 1.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
CC APHTHOVIRUS.
[1]
RN RN SEQUENCE FROM N.A.
RC STRAIN-PERV;
RX MEDLINE; 96332508.
RA WUTZ G., AUER H., NOWOTNY N., SKERN T., KUECHLER E.;
RT "Equine rhinovirus serotypes 1 and 2: relationship to each other and
RL rt to aphthoviruses and cardioviruses." ;
```

```

RESULT 11
ID OC08906 PRELIMINARY; PRT: 1300 AA.
AC AC08906;
DT 01-JUL-1997 (TREMBLER. 04, CREATED)
DT 01-JUL-1997 (TREMBLER. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLER. 08, LAST ANNOTATION UPDATE)
DE REVERSE TRANSCRIPTASE (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIROGONATHI; MURIDAE; MURINAE; MUS.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-ORLEANS REELER; TISSUE-LIVER;
RX MEDLINE; 96414308.
RA TAKAHARA T., OHSUMI T., KUROMITSU J., SHIBATA K., SASAKI N.,
RA OKAZAKI Y., SHIBATA H., SATO S., YOSHIZAKI A., KUSAKABE M.,
RA MURAMATSU M., UKI M., OKUDA K., HAYASHIZAKI Y.
RT "dysfunction of the Orleans reeler gene arising from exon skipping
RT due to transposition of a full-length copy of an active L1 sequence
RT into the skipped exon";
RL HUM. MOL. GENET. 5:989-993(1996).
DR EMBL; D84391; D1021252; -.
DR PFAM; PF00078; rvt; 1.
DR PFAM; PF01260; AP_endonuclease1; 1.
KW RNA-DIRECTED DNA POLYMERASE.
FT NON_TER 1
FT SEQUENCE 1300 AA; 152068 MW; 3656A0F1 CRC32;
SQ
Query Match 4.2%; Score 106; DB 11; Length 1300;
Best Local Similarity 27.1%; Pred. No. 3.55e-01;
Matches 23; Conservative 22; Mismatches 38; Indels 2; Gaps

```

```
QY   96 NSLTNNRIVALTSGKYNFPPEVYSWLRTAGTSAPFSDSILSRFDVSAAAFYTSSKRAIA 155

Db    150 FTKETLVKLKAHIAPHITIIIGDENT 174
      : | ::|::||::|::
QY   156 LEHVLSNRKSTDDYQTILDVVDFS 180

RESULT 12
ID     O39818
AC     Q39818; PRELIMINARY; PRT: 2247 AA.
DT     01-JAN-1998 (TREMBUREL. 05, CREATED)
DT     01-JAN-1998 (TREMBUREL. 05, LAST SEQUENCE UPDATE)
DE     01-NOV-1998 (TREMBUREL. 08, LAST ANNOTATION UPDATE)
DS     POLYPROTEIN.
OS     EQUINE RHinovIRUS 1.
SC     VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE.
CN     [1]
RN     SEQUENCE FROM N.A.
RX     MEDLINE: 96165505.
RA     LI F., BROWNING G.F., STUDDERT M.J., CRABB B.S.;
RT     "Equine rhinovirus 1 is more closely related to foot-and-mouth
RT     disease virus than to other picornaviruses.";
RL     PROC. NATL. ACAD. SCI. U.S.A. 93:990-995(1996).
DR     ENBL; L43052; G2231134; -.
DR     PFAM; PF00073; rhv; 1.
DR     PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR     PFAM; PF00910; RNA_helicase; 1.
KW     POLYPROTEIN.
SQ     SEQUENCE 2247 AA; 249301 MW; 2DD1F630 CRC32;

Query Match          4.2%; Score 107; DB 14; Length 2247;
Best Local Similarity 23.4%; Pred. No. 2.63e-01;
Matches             22; Conservative 33; Mismatches 32; Indels 7; Gaps 7

Db    1041 TYPSPFKMAKWLFLATLAVAAMRTKPVVVMYLADI-FGLE-VEDTGFFFSYFEQEKLP 1098
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
QY   142 SYAAFYTTSSRAIALEHVKLSNRKSTDYQITLDVV-FDSLLEDVGATGPRGTY-ESVEQ 199
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::

Db    1099 YMKTIPIKISDLVTDAATAAQAIPK-GVISFVSS 1131
```





```

RP  SEQUENCE FROM N.A.
RC  STRAIN=TRINIDAD DONKEY;
RA  KINNEY R.;
RL  SUBMITTED (AUG-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
DR  EMBL; L01442; G3201595; -.
DR  PFAM; PF00978; RNA_dep_RNapol2; 2.
KW  POLYPROTEIN.
FT  CHAIN 1 535 NONSTRUCTURAL PROTEIN NSP1.
FT  CHAIN 536 1329 NONSTRUCTURAL PROTEIN NSP2.
FT  CHAIN 1330 1886 NONSTRUCTURAL PROTEIN NSP3.
FT  CHAIN 1887 2493 NONSTRUCTURAL PROTEIN NSP4.
SQ  SEQUENCE 2493 AA; 277915 MW; A6B7B200 CRC32;

Query Match 4.2%; Score 106; DB 14; Length 2493;
Best Local Similarity 24.2%; Pred. No. 3.55e-01;
Matches 23; Conservative 26; Mismatches 39; Indels 7; Gaps 7;

Db 1125 LRNYDPRINLPVNNRPLHALVLHNEHPOSDFSSFVSKL-KGRTVLVY-GEKLSVPGKM 1182
QY 62 LRVFDPSTGAL-VDSK-SYAFSTN-NDTTSAAFYSEFMNSLTNNRIVAILTSGKVNFPPEV 118
Db 1183 VDWLSDRPEATFRARLDLIGPDVVKYDIIFVNVN 1217
QY 119 VSWLRTAGTSAPFSDSILS-RFDVS-YAAPTSSK 151

```

Search completed: Wed Sep 1 16:02:23 1999  
Job time : 64 secs.

\*\*\*\*\*  
 W P S R E L E A  
 \*\*\*\*\*  
 (TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Wed Sep 1 16:06:41 1999; MasPar time 4.33 Seconds  
 Tabular output not generated.  
 117.961 Million cell updates/sec

Title: >PCT-US99-13024-2  
 Description: (1-24) from PCTUS9913024.pep (2 of 12)  
 Perfect Score: 176  
 Sequence: 1 MEKFMARFGQGVQTPFLSESNSV 24  
 Scoring table: PAM 150  
 Gap 11

Searched: 170751 seqs, 21266608 residues  
 Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-gensseq35  
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21 22:part22 23:part23  
 24:part24 25:part25 26:part26 27:part27 28:part28  
 29:part29 30:part30 31:part31 32:part32 33:part33  
 34:part34 35:part35 36:part36 37:part37 38:part38  
 39:part39

Statistics: Mean 21.293; Variance 79.022; scale 0.269

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	74	42.0	65	19 R97371	Phage T4 ORF gene pr	8.97e+00
2	65	36.9	591	37 W72912	Mycobacterium tubercu	5.69e+01
3	65	36.9	621	29 W55645	H. pylori ORF 06ep106	5.69e+01
4	65	36.9	637	30 W55739	H. pylori ORF 06ep106	5.69e+01
5	65	36.9	645	37 W73030	Helicobacter pylori 7	5.69e+01
6	65	36.9	652	37 W73029	Helicobacter pylori 7	5.69e+01
7	64	36.4	358	35 W44826	Pseudomonas fluoresce	6.96e+01
8	64	36.4	358	35 W44826	Pseudomonas fluoresce	6.96e+01
9	63	35.8	364	33 W62047	Streptococcus pneumon	8.50e+01
10	63	35.8	389	2 P70668	D-alanine racemase.	8.50e+01
11	63	35.8	409	9 W48060	Sequence of protease	8.50e+01
12	63	35.8	461	30 W39928	Human CDP-diacylglyce	8.50e+01
13	63	35.8	740	33 W62045	Streptococcus pneumon	8.50e+01
14	62	35.2	146	29 W55354	H. pylori ORF 01cp117	1.04e+02
15	62	35.2	269	29 W55318	H. pylori ORF 07ep119	1.04e+02
16	62	35.2	479	29 W55555	H. pylori ORF 02ge101	1.04e+02

17	62	35.2	479	29 W55556	H. pylori ORF 02ge101	1.04e+02
18	62	35.2	486	29 W55502	H. pylori ORF 02gp207	1.04e+02
19	61	34.7	43	8 R40095	Hib OMP P1-P2 hybrid	1.26e+02
20	61	34.7	357	29 W55533	H. pylori ORF hp6e203	1.26e+02
21	61	34.7	382	29 W55369	Breast tumor kinase,	1.26e+02
22	61	34.7	451	12 R63088	H. pylori ORF 04gp112	1.26e+02
23	61	34.7	479	29 W55693	H. pylori ORF 09cp107	1.26e+02
24	61	34.7	479	29 W55570	H. pylori ORF 06cp306	1.26e+02
25	61	34.7	479	29 W55569	H. pylori ORF 06cp306	1.26e+02
26	61	34.7	479	29 W55702	H. pylori ORF 09cp107	1.26e+02
27	61	34.7	510	26 W34561	Pyrococcus furiosus g	1.26e+02
28	61	34.7	510	11 R58801	Hyperthermostable be	1.26e+02
29	61	34.7	510	17 R88094	Hyperthermostable bet	1.26e+02
30	61	34.7	510	36 W49865	Pyrococcus VC1-7G1 gl	1.26e+02
31	60	34.1	85	29 W55416	H. pylori ORF hp1p105	1.54e+02
32	60	34.1	130	35 W80717	S. pneumoniae protein	1.54e+02
33	60	34.1	313	37 W75159	Human secreted protein	1.54e+02
34	60	34.1	390	23 W21678	Haemophilus influenza	1.54e+02
35	60	34.1	435	29 W55642	Human calcium channel	1.54e+02
36	60	34.1	451	5 R27642	Calcium ion channel a	1.54e+02
37	60	34.1	2262	31 W56737	Leader peptide sequen	1.87e+02
38	59	33.5	38	35 W78736	Streptococcus pneumon	1.87e+02
39	59	33.5	241	34 W38602	Glycerol dehydrogenas	1.87e+02
40	59	33.5	380	35 W69360	Wzy (Rfc) protein inv	1.87e+02
41	59	33.5	438	28 W37355	Human cytomegalovirus	1.87e+02
42	59	33.5	946	26 W27085	BamW coat protein.	1.87e+02
43	59	33.5	1202	8 R44507	GAP protein Irai.	1.87e+02
44	59	33.5	2938	11 R59223	Saccharomyces cerevis	1.87e+02
45	59	33.5	2942	25 W22050		

## ALIGNMENTS

RESULT 1  
 ID R97371 standard; Protein: 65 AA.

AC R97371;  
 DT 07-JAN-1997 (first entry)  
 DE Phage T4 ORF gene product, vgp34.  
 KW Phage T4; tail fibre protein; nanotechnology; nano-structure;  
 KW filter; molecular sieve.  
 OS Bacteriophage T4.  
 PN W09611947-A1.  
 PD 25-APR-1996.  
 PF 13-OCT-1995; U13023.  
 PR 13-OCT-1994; US-322760.  
 PA (GOLD/) GOLDBERG E B.  
 PI Goldberg EB;  
 DR WPI: 96-221942/22.  
 DR N-PSDB; T29053.  
 PT New proteins derived from T4 phage tail fibre proteins - that can  
 PT self assemble into nano-structure(s), useful as filters etc, also  
 PT corresponding DNA  
 PS Claim 7; Fig 7; 83pp; English.  
 CC A protein (R97371) of unspecified function is the product of open  
 CC reading frame x of the tail fibre protein gene region (see also  
 CC T29053) of phage T4. This gene region also includes open reading  
 CC frames for tail fibre proteins (see also R97370 and R97372-74).  
 CC Tail fibre proteins (native or modified) can be produced in large  
 CC quantities in microbial cells and used as building blocks of strong,  
 CC stable nanostructures.  
 SQ Sequence 65 AA;

Query Match 42.0%; Score 74; DB 19; Length 65;  
 Best Local Similarity 58.3%; Pred. No. 8.97e+00;  
 Matches 14; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

DB 1 mekfmaefgqgvqtpflsesns 24  
 QY 1 MEKFMARFGQGVQTPFLSESNSV 24

RESULT 2  
 ID W72912 standard; Protein: 591 AA.

AC W72912;  
 DT 21-JAN-1999 (first entry)  
 DE Mycobacterium tuberculosis antigen RDI-ORF2.  
 KW Mycobacterium tuberculosis; antigen; vaccine; immunological;  
 OS immunogen; infection.  
 PN WO9844119-A1.  
 PD 08-OCT-1998.  
 PF 01-APR-1998; DK0132.  
 PR 05-JAN-1998; US-070488.  
 PR 02-APR-1997; DK-000376.  
 PR 18-APR-1997; US-044624.  
 PR 10-NOV-1997; DK-001277.  
 PA (STAT-) STATENS SERUM INST.  
 PI Andersen P, Florio W, Nielsen R, Oettinger T, Rasmussen PB,  
 PI Rosenkrands I, Wellingh K;  
 DR WPI: 98-542705/45.  
 DR N-PSDB: V63936.  
 PT New isolated mycobacteria polypeptides and nucleic acids - used for  
 PT developing products for the diagnosis of or vaccination against  
 PT mycobacterial infections, particularly tuberculosis  
 PS Claim 1: Page 188-190; 163pp; English.  
 CC The present sequence represents a Mycobacterium tuberculosis protein.  
 CC Products from the present invention, which describes protein fragments  
 CC and nucleic acid fragments derived from M.tuberculosis, can be used in  
 CC the detection of and prevention of mycobacterial infections. In  
 CC particular, the proteins and nucleic acids can be used for the diagnosis  
 CC of or vaccination against tuberculosis caused by M. tuberculosis,  
 CC M. africanum or M. bovis.  
 CC Sequence 591 AA;  
 SQ

Query Match 36.9%; Score 65; DB 37; Length 591;  
 Best Local Similarity 55.0%; Pred. No. 5.69e+01;  
 Matches 11; Conservative 2; Mismatches 5; Indels 2; Gaps 2;

Db 522 mdkfvgaafgsg-aptmfsls 540

QY 1 MEKFM-AEFGQGVQTPFLS 19

## RESULT 3

ID W55645 standard; Protein: 621 AA.  
 AC W55645;  
 DT 24-JUN-1998 (first entry)  
 DE H. pylori ORF 06ep10615\_961562\_f2\_41 cell envelope OMP.  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacteria; life cycle; activator;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.  
 OS Helicobacter pylori.  
 PN W09737044-A1.  
 PD 09-OCT-1997.  
 PF 27-MAR-1997; U05223.  
 PR 06-DEC-1996; US-761318.  
 PR 29-MAR-1996; US-625811.  
 PR 02-APR-1996; US-758731.  
 PR 25-OCT-1996; US-736905.  
 PR 28-OCT-1996; US-738859.  
 PA (ASTR ) ASTRA AB.  
 PI Alm RA, Smith D;  
 DR WPI: 97-503122/46.  
 DR N-PSDB: V25054.  
 PT Helicobacter pylori nucleic acid sequences and encoded  
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
 PT infection and for diagnosis of H. pylori infection  
 PS Claims 14, 80; Pages 850-851; 1145pp; English.  
 CC This is a H. pylori cell envelope outer membrane protein having a C-  
 CC terminal tyrosine cluster motif. The protein may be used in a vaccine to  
 CC prevent or treat H. pylori infection or to identify H. pylori polypeptide  
 CC binding compounds, useful as potential H. pylori life cycle activators or  
 CC inhibitors. The DNA and probes derived from it may be used for the  
 CC identification of H. pylori in a sample and the diagnosis of H. pylori  
 CC infection. Nucleic acid sequences complementary to the DNA act as  
 CC antisense sequences and can be used to prevent the translation of

H. pylori mRNA. Antibodies against the protein can be used in  
 CC immunoassays to evaluate the abundance and distribution of H. pylori-  
 CC specific antigens. The genomic sequence of H. pylori (ATCC 55679) was  
 CC determined from overlapping contigs generated by mechanically shearing  
 CC the bacterial DNA. The sequences were analysed for ORF of at least 180  
 CC nucleotides, and the predicted coding regions defined by computer  
 CC evaluation. To identify likely H. pylori antigens for vaccine  
 CC development, the amino acid sequences predicted from various ORF were  
 CC analysed for significant homology to other known or exported membrane  
 CC proteins. Having identified and determined the sequences of interest,  
 CC particular regions can be isolated from H. pylori by PCR amplification  
 CC for recombinant polypeptide production, e.g. in E. coli hosts.  
 SQ Sequence 621 AA;

Query Match 36.9%; Score 65; DB 29; Length 621;  
 Best Local Similarity 22.2%; Pred. No. 5.69e+01;  
 Matches 4; Conservative 10; Mismatches 4; Indels 0; Gaps 0;

Db 475 dynhykssfnssd1 492

QY 7 EFGQGVQTPFLSESNSV 24

## RESULT 4

ID W55739 standard; Protein: 637 AA.  
 AC W55739;  
 DT 13-JUL-1998 (first entry)  
 DE H. pylori ORF 06ep10615\_961562\_f1\_15 cell envelope OMP.  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacteria; life cycle; activator;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;  
 KW cell envelope outer membrane protein; OMP; tyrosine cluster motif.  
 OS Helicobacter pylori.  
 PN W09737044-A1.  
 PD 09-OCT-1997.  
 PF 27-MAR-1997; U05223.  
 PR 06-DEC-1996; US-761318.  
 PR 29-MAR-1996; US-625811.  
 PR 02-APR-1996; US-758731.  
 PR 25-OCT-1996; US-736905.  
 PR 28-OCT-1996; US-738859.  
 PA (ASTR ) ASTRA AB.  
 PI Alm RA, Smith D;  
 DR WPI: 97-503122/46.  
 DR N-PSDB: V25148.  
 PT Helicobacter pylori nucleic acid sequences and encoded  
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
 PT infection and for diagnosis of H. pylori infection  
 PS Claims 14, 80; Page 1108,1109; 1145pp; English.  
 CC This sequence represents a Helicobacter pylori cell envelope outer  
 CC membrane protein having a C-terminal tyrosine cluster motif.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors. The  
 CC DNA and probes derived from it may be used for the identification of  
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic  
 CC acid sequences complementary to the DNA act as antisense sequences and  
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies  
 CC against the protein can be used in immunoassays to evaluate the abundance  
 CC and distribution of H. pylori-specific antigens. The genomic sequence of  
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated  
 CC by mechanically shearing the bacterial DNA. The sequences were analysed  
 CC for ORF of at least 180 nucleotides, and the predicted coding regions  
 CC defined by computer evaluation. To identify likely H. pylori antigens for  
 CC vaccine development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or exported  
 CC membrane proteins. Having identified and determined the sequences of  
 CC interest, particular regions can be isolated from H. pylori by PCR  
 CC amplification for recombinant polypeptide production, e.g. in E. coli  
 CC hosts.  
 SQ Sequence 637 AA;

Query Match 36.9%; Score 65; DB 30; Length 637;

```

Best Local Similarity 22.2%; Pred. No. 5.69e+01;
Matches 4; Conservative 10; Mismatches 4; Indels 0; Gaps 0;

Db 491 dnhgviksfsssd1 508
QY 7 EFGQGVQVPFLSESNSV 24
:::||||:|:|:|

RESULT 5
ID W73030 standard; Protein; 645 AA.
AC W73030;
DT 02-FEB-1999 (first entry)
DE Helicobacter pylori 76 kDa polypeptide GHPO 185.
KW GHPO 185; infection; gastritis; ulcer; vaccine; diagnosis;
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT Peptide 1..19 /label= Sig_peptide
FT Protein 20..645 /label= Mat_protein
FT W09843479-A1.
PN 08-OCT-1998.
PD 08-OCT-1998.
PF 31-MAR-1998; U06421.
PR 01-APR-1997; US-834666.
PR 01-APR-1997; US-831310.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Al-Garawi A, Kleanthous H, Lissolo L, Miller C, Tomb J;
WPI: 98-568251/48.
DR N-PSDB; V07919.
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastroduodenal diseases
PS Claim 9: Page 123-125; 184pp; English.
CC This is the amino acid sequence of a 76 kDa Helicobacter pylori
CC polypeptide designated GHPO 185. It was deduced from an isolated
CC genomic DNA sequence (see V07919). The invention provides a family
CC of 76 kDa Helicobacter polypeptides (see W73022-32), as well as a
CC 32 kDa polypeptide (see W73034) and a 50 kDa polypeptide (see
CC W73035), and also polynucleotides (see V72001, V07912-21 and V07963-64)
CC encoding them, expression cassettes, and methods for producing the
CC unprocessed or mature polypeptides in host cells. The polypeptides
CC can be used in vaccination methods to prevent or treat Hb infection
CC in a mammal. Methods and products of the invention allow treatment
CC and prevention of gastroduodenal diseases associated with Hb
CC infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. Detection
CC and diagnostic methods are also provided. GHPO 190 was
CC demonstrated to be a protective antigen.
SQ Sequence 645 AA;

Query Match 36.9%; Score 65; DB 37; Length 645;
Best Local Similarity 22.2%; Pred. No. 5.69e+01;
Matches 4; Conservative 10; Mismatches 4; Indels 0; Gaps 0;

Db 506 dnhgviksfsssd1 523
QY 7 EFGQGVQVPFLSESNSV 24
:::||||:|:|:|

RESULT 7
ID W44826 standard; Protein; 358 AA.
AC W44826;
DT 23-NOV-1998 (first entry)
DE Pseudomonas fluorescens strain P.J.874 HPD protein.
DE Transgenic plant; resistance; herbicide; chimaeric gene; tolerance; PCR;
KW hydroxyphenylpyruvate dioxygenase; isoxaflutole; sulcotrione; primer;
KW amplification; tobacco; cotton; rape; soya; sugar beet; cereal; loxynil;
KW nitrilase; dihalohydroxybenzotrile; bromoxynil; nematode; insecticide;
KW glyphosate oxidoreductase; Bacillus thuringiensis; fungus.
OS Pseudomonas fluorescens.
PN W09802362-A2.
PD 22-JAN-1998.
PF 10-JUL-1997; F01256.
PF 16-JUL-1996; FR-009137.
PA (RHON) RHONE-POULENC AGROCHIMIE.
PI Derose R, Pallett K, Pelissier B, Sailland A;
WPI: 98-110608/10.
DR N-PSDB; V19280.
PT Chimaeric gene imparting resistance to two or more herbicides - such
PT as hydroxyphenylpyruvate dioxygenase inhibitors, also related
PT vectors and transformed plants
PS Disclosure; Fig 1; 49pp; French.
CC The invention relates to the generation of transgenic plants resistant
CC to preferably at least 2 herbicides by introducing a chimaeric gene
CC comprising at least 2 gene subunits where one sequence encodes
CC hydroxyphenylpyruvate dioxygenase (HPD), an enzyme that promotes
CC tolerance to herbicides such as isoxaflutole or sulcotrione. This
CC sequence represents the HPD protein from Pseudomonas fluorescens strain
CC P.J.874. The sequence was used to derive the degenerate coding sequence
CC (V19280). This in turn was used to design the degenerate primers
CC (V19270-V19274) for the isolation of the HPD gene from P. fluorescens
CC strain A32 (V19269). The plants, e.g. tobacco, cotton, rape, soya,

```

```

Query Match      35.88;      Score 63;  DB 2;  Length 389;
Best Local Similarity 50.0%;  Pred. No. 8.50e+01;
Matches 6;  Conservative 2;  Mismatches 4;  Indels 0;  Gaps 0;

Db 313 mdqfvmvldgqy 324
QY 1 MERFMAEFQGQY 12
    | | | | |

```

RESULT	9	
ID	W62047	standard; Protein: 364 AA.
AC	W62047;	
DT	01-OCT-1998	(first entry)
DE	Streptococcus pneumoniae	spo/rel distal terminal portion.
DE	Streptococcus pneumoniae:	spo/rel family; screening;
KW	antibacterial; otitis media;	conjunctivitis; pneumonia; bacteraemia;
KW	meningitis.	
KW	Streptococcus pneumoniae.	
OS	Streptococcus pneumoniae.	
PN	EP-849362-A2.	

CC polypeptide with a drug candidate (preferably a pool of compounds  
CC from combinatorial library expression), and measuring CDS activity  
CC with and without different concentrations of the drug candidate.  
CC Modulator compounds can be used e.g. as inhibitors of the  
CC inflammatory cascade following hypoxia and reoxygenation, as in  
CC cases of sepsis, trauma or adult respiratory distress syndrome.  
CC CDS is also used to raise specific antibodies for therapeutic or  
CC diagnostic use.  
SQ Sequence 461 AA;

Query Match 35.8%; Score 63; DB 30; Length 461;  
Best Local Similarity 40.0%; Pred. No. 8.50e+01;  
Matches 8; Conservative 7; Mismatches 4; Indels 1; Gaps 1;

Db 405 fmatlvhgyi-tsfigrnpn 423  
||| : ||| : ||| : ||| :  
QY 4 FMAEFGGYVQTPTFLSESNS 23

RESULT 13  
ID W62045 standard; Protein: 740 AA.  
AC W62045;  
DT 01-OCT-1998 (first entry)  
DE Streptococcus pneumoniae spo/rel protein sequence.  
KW Streptococcus pneumoniae; spo/rel; spo/relA family; screening;  
KW antibacterial; otitis media; conjunctivitis; pneumonia; bacteraemia;  
KW meningitis.  
OS Streptococcus pneumoniae.  
PN EP-849362-A2.  
PS 24-JUN-1998.  
PD 24-JUN-1998.  
PF 23-OCT-1997; 308473.  
PR 24-OCT-1996; US-029049.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PI Gentry DR;  
DR WPI: 98-324673/29.  
DR N-PSDB: V37946.  
PT New Streptococcus spo/relA polypeptide(s) - useful in treatment of  
PT otitis media, conjunctivitis, pneumonia, bacteremia and especially  
PT meningitis  
PS Claim 11; Page 26-28; 40pp; English.  
CC The present sequence represents a spo/rel protein (spo/relA family) from  
CC Streptococcus pneumoniae. The protein is of the spo/relA family, which,  
CC in Escherichia coli, is involved in the stringent response to nutrient  
CC limitation and regulate the accumulation of (p)ppGpp which is involved  
CC in the regulation of gene expression and other cellular processes.  
CC Spo/rel proteins have a homology to e.g. S. equisimilis rel protein.  
CC Spo/rel proteins can be used: (a) in the treatment of an individual in  
CC need of spo/rel protein, by administering to the protein to the patient;  
CC (b) in the treatment of an individual having need to inhibit spo/rel  
CC protein, by administering an antagonist which inhibits the activity of  
CC the protein; and (c) for inducing an immunological response by  
CC inoculating the mammal with spo/rel protein, or a fragment or variant  
CC of it, adequate to produce antibody and/or T cell immune response to  
CC protect the animal from disease. Conditions which may be treated  
CC include otitis media, conjunctivitis, pneumonia, bacteraemia, meningitis  
CC sinusitis, pleural empyema, endocarditis and especially meningitis and  
CC compositions may be used as antibacterials.  
SQ Sequence 740 AA;

Query Match 35.8%; Score 63; DB 33; Length 740;  
Best Local Similarity 44.4%; Pred. No. 8.50e+01;  
Matches 8; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 495 emlmaeqfengyankfm 512  
||:||||: |||| : ||| :  
QY 2 EKFAEFGQ-GYVQTPTFL 18

RESULT 14  
ID W55354 standard; Protein: 146 AA.  
AC W55354;  
DT 17-JUN-1998 (first entry)  
DE H. pylori ORF 01cpl1710orf5 protein.

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacteria; life cycle; activator;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.  
 OS Helicobacter pylori.  
 PN W09737044-A1.  
 PD 09-OCT-1997. U05223.  
 PF 27-MAR-1997. US-761318.  
 PR 06-DEC-1996; US-761318.  
 PR 29-MAR-1996; US-625811.  
 PR 02-APR-1996; US-758731.  
 PR 25-OCT-1996; US-736905.  
 PR 28-OCT-1996; US-738859.  
 PA (ASTR ) ASTRA AB.  
 PI Alm RA, Smith D;  
 DR WPI; 97-503122/46.  
 DR N-PSDB; V24764.  
 PT Helicobacter pylori nucleic acid sequences and encoded  
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
 PT infection and for diagnosis of H. pylori infection  
 PS Claim 14; Page 576-577; 1145pp; English.  
 CC This sequence is a H. pylori protein of unspecified function.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors. The  
 CC DNA and probes derived from it may be used for the identification of  
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic  
 CC acid sequences complementary to the DNA act as antisense sequences and  
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies  
 CC against the protein can be used in immunoassays to evaluate the abundance  
 CC and distribution of H. pylori-specific antigens. The genomic sequence of  
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated  
 CC by mechanically shearing the bacterial DNA. The sequences were analysed  
 CC for ORF of at least 180 nucleotides, and the predicted coding regions  
 CC defined by computer evaluation. To identify likely H. pylori antigens for  
 CC vaccine development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or exported  
 CC membrane proteins. Having identified and determined the sequences of  
 CC interest, particular regions can be isolated from H. pylori by PCR  
 CC amplification for recombinant polypeptide production, e.g. in E. coli  
 CC hosts.  
 SQ Sequence 146 AA;  
 Query Match 35.2%; Score 62; DB 29; Length 146;  
 Best Local Similarity 45.0%; Pred. No. 1.04e+02;  
 Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 Db 35 etftvavqgniyadflsks 54  
 | | : : : : :  
 QY 2 EKFWAEFGQGVQTPFLSES 21  
 RESULT 15  
 ID W55518 standard; Protein; 269 AA.  
 AC W55518:  
 DT 02-JUL-1998 (first entry)  
 DE H. pylori ORF 07ep11916\_5913592\_f3\_18 cell envelope OMP.  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacteria; life cycle; activator;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.  
 OS Helicobacter pylori.  
 PN W09737044-A1.  
 PD 09-OCT-1997. U05223.  
 PF 27-MAR-1997. US-761318.  
 PR 06-DEC-1996; US-761318.  
 PR 29-MAR-1996; US-625811.  
 PR 02-APR-1996; US-758731.  
 PR 25-OCT-1996; US-736905.  
 PR 28-OCT-1996; US-738859.  
 PA (ASTR ) ASTRA AB.  
 PI Alm RA, Smith D;  
 DR WPI; 97-503122/46.  
 DR N-PSDB; V24927.  
 PT Helicobacter pylori nucleic acid sequences and encoded

PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
 PT infection and for diagnosis of H. pylori infection  
 PS Claims 14,80; Page 721-722; 1145pp; English.  
 CC This sequence is a H. pylori cell envelope outer membrane protein which  
 CC is stated to have a terminal Phe residue.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors. The  
 CC DNA and probes derived from it may be used for the identification of  
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic  
 CC acid sequences complementary to the DNA act as antisense sequences and  
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies  
 CC against the protein can be used in immunoassays to evaluate the abundance  
 CC and distribution of H. pylori-specific antigens. The genomic sequence of  
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated  
 CC by mechanically shearing the bacterial DNA. The sequences were analysed  
 CC for ORF of at least 180 nucleotides, and the predicted coding regions  
 CC defined by computer evaluation. To identify likely H. pylori antigens for  
 CC vaccine development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or exported  
 CC membrane proteins. Having identified and determined the sequences of  
 CC interest, particular regions can be isolated from H. pylori by PCR  
 CC amplification for recombinant polypeptide production, e.g. in E. coli  
 CC hosts.  
 SQ Sequence 269 AA;  
 Query Match 35.2%; Score 62; DB 29; Length 269;  
 Best Local Similarity 45.0%; Pred. No. 1.04e+02;  
 Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 Db 184 etftvavqgniyadflsks 203  
 | | : : : : :  
 QY 2 EKFWAEFGQGVQTPFLSES 21

Search completed: Wed Sep 1 16:07:08 1999  
 Job time : 27 secs.

\*\*\*\*\*

MPERCH\_PP protein - protein database search, using Smith-Waterman algorithm

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Run on: Wed Sep 1 16:06:10 1999; MasPar time 4.18 Seconds  
Tabular output not generated. 229.968 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (1-24) from PCTUS9913024.pep (2 of 12)  
Perfect Score: 176  
Sequence: 1 MEKFMAEFGQGYVOTPLSESNSV 24

Scoring table: PAM 150  
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 30.660; Variance 51.855; scale 0.591

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	77	43.8	850	2	F71000 hypothetical protein	1.61e-01
2	73	41.5	2301	2	T02323 hypothetical protein	7.07e-01
3	68	38.6	554	2	S41574 DNA ligase (ATP) (EC	4.22e+00
4	68	38.6	809	1	Q0B834 BBLF4 protein - human	4.22e+00
5	68	38.6	1224	2	A25884 DNA-directed RNA poly	4.22e+00
6	67	38.1	236	2	JQ0061 hypothetical 25K prot	5.97e+00
7	66	37.5	56	2	S35168 cytochrome P450 (clon	8.42e+00
8	66	37.5	70	2	E64757 probable membrane pro	8.42e+00
9	66	37.5	112	1	GMPT4 gene 33 protein - pha	8.42e+00
10	66	37.5	336	2	B71121 hypothetical protein	8.42e+00
11	66	37.5	337	2	S21209 4-hydroxyphenylpyruva	8.42e+00
12	66	37.5	508	2	I64032 hypothetical protein	8.42e+00
13	66	37.5	518	2	H64775 probable membrane pro	8.42e+00
14	66	37.5	853	2	H70470 hypothetical protein	8.42e+00
15	66	37.5	1191	2	S65068 DNA-directed RNA poly	8.42e+00
16	65	36.9	277	2	S32571 malD protein - Strept	1.18e+01
17	65	36.9	591	2	E70802 hypothetical protein	1.18e+01
18	65	36.9	608	2	B64610 outer membrane protei	1.18e+01
19	65	36.9	629	2	E64610 outer membrane protei	1.18e+01
20	65	36.9	638	2	B71904 probable outer membra	1.18e+01
21	65	36.9	651	2	E71904 probable outer membra	1.18e+01
22	64	36.4	182	2	I64029 hypothetical protein	1.66e+01
23	64	36.4	273	2	D70428 hypothetical protein	1.66e+01

24	64	36.4	420	1	TVUTG4	phosphoglycerate kina	1.66e+01
25	64	36.4	440	1	TVUTGB	phosphoglycerate kina	1.66e+01
26	64	36.4	440	1	KIUTGG	phosphoglycerate kina	1.66e+01
27	64	36.4	505	1	TVUT2B	phosphoglycerate kina	1.66e+01
28	64	36.4	508	1	TVUT4B	phosphoglycerate kina	1.66e+01
29	64	36.4	509	1	A45593	phosphoglycerate kina	1.66e+01
30	64	36.4	656	2	S61097	hypothetical protein	1.66e+01
31	64	36.4	980	2	T00045	celldextrin phosphor	1.66e+01
32	64	36.4	1188	2	S30229	DNA-directed RNA poly	1.66e+01
33	63	35.8	175	2	G71085	hypothetical protein	2.31e+01
34	63	35.8	278	2	A70034	maltoedextrin transpor	2.31e+01
35	63	35.8	389	1	JS0043	alanine racemase (EC	2.31e+01
36	63	35.8	389	1	S36553	kpsD protein - Escher	2.31e+01
37	63	35.8	417	1	TVCRCG	phosphoglycerate kina	2.31e+01
38	63	35.8	455	2	E71569	probable acyltransfer	2.31e+01
39	63	35.8	455	1	TVCRCG	phosphoglycerate kina	2.31e+01
40	63	35.8	603	1	W1WL8	E1 protein - human pa	2.31e+01
41	63	35.8	645	2	F70825	probable PPE protein	2.31e+01
42	63	35.8	686	2	S48467	hypothetical protein	2.31e+01
43	63	35.8	730	2	S54625	probable membrane pro	2.31e+01
44	63	35.8	1215	2	S60304	hypothetical protein	2.31e+01
45	62	35.2	402	2	S23860	chloramphenicol resis	3.21e+01

ALIGNMENTS

RESULT 1

ENTRY

TITLE

ORGANISM

DATE

ACCESSIONS

REFERENCE

#authors

#type complete

#formal\_name Pyrococcus horikoshii

14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change

F71000

hypothetical protein PH1302 - Pyrococcus horikoshii

14-Aug-1998

F71000

Kawarabayashi, Y.; Savada, M.; Horikawa, H.; Haikawa, Y.;

Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;

Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;

Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;

Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,

A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;

Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.

DNA Res. (1998) 5:55-76

Complete sequence and gene organization of the genome of a

hyper-thermophilic archaeobacterium, Pyrococcus horikoshii

OT3.

#cross-references MOID:98344137

#accession F71000

#status preliminary; nucleic acid sequence not shown;

translation not shown

#molecule\_type DNA

#residues 1-850 #label KAW

#cross-references GB:AP000006; MID:g2336133; PID:d1031349; PID:g3557723

#experimental\_source strain OT3

#note this accession replaces an interim accession for a

sequence replaced by GenBank

GENETICS

#gene PH1302

#length 860 #molecular-weight 95454 #checksum 3711

Query Match 43.8%; Score 77; DB 2; Length 860;

Best Local Similarity 29.4%; Pred. No. 1.61e-01;

Matches 5; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

Db 1 MSKFLDPSNATIRVAF 17

| | | | | : : : : : |

QY 1 MEKFMAEFGQGYVQTPF 17

| | | | | : : : : : |

RESULT 2

ENTRY

TITLE

ORGANISM

#type complete

hypothetical protein F13P17.19 - Arabidopsis thaliana

#formal\_name Arabidopsis thaliana #common\_name mouse-ear

cross



```

DATE      05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
ACCESSIONS T02323
REFERENCE   Z14168
#authors    Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.;
            Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.;
            Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter,
            J.C.
#submission submitted to the EMBL Data Library, July 1998
#description Arabidopsis thaliana chromosome II BAC F13P17 genomic
            sequence.
#accession  T02323
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues   1-2301 #label ROU
#cross-references EMBL:AC004481; NID:g3337347; PID:g3337366
GENETICS
#map_position 2
#introns      155/3; 297/1; 525/2; 550/1; 609/1; 767/3; 911/1; 1139/2;
            1238/3; 1412/1; 1498/1; 1528/3; 1598/3; 1732/3; 1791/2;
            1822/3; 1841/3; 1884/3; 1907/3; 1945/3; 1999/3; 2033/3;
            2060/2; 2107/2
#note       F13P17.19
SUMMARY     #length 2301 #molecular-weight 253398 #checksum 4265
Query Match 41.5%; Score 73; DB 2; Length 2301;
Best Local Similarity 50.0%; Pred. No. 7.07e-01;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
Db 977 NFLGRFGAGYVSDTFL 992
   ||| ||| |||
Qy 3 KFMFAEFGQGYVOTPF 18

RESULT      3
ENTRY       S41974 #type complete
TITLE       DNA ligase (ATP) (EC 6.5.1.1) - fowlpox virus
ORGANISM    #formal_name fowlpox virus
DATE        06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
ACCESSIONS S41974
REFERENCE   S41971
#authors    Skinner, M.A.; Moore, J.B.; Binns, M.M.; Boursnell, M.E.
#submission submitted to the EMBL Data Library, February 1994
#description Deletion of fowlpox virus homologues of vaccinia virus genes
            between the 3b-hydroxysteroid dehydrogenase (A44L) and DNA
            ligase (A50R) genes.
#accession  S41974
#status     preliminary
#molecule_type DNA
#residues   1-564 #label SK1
#cross-references EMBL:229716; NID:g453598; PID:g453602
CLASSIFICATION #superfamily vaccinia virus DNA ligase
KEYWORDS     ligase; phosphoprotein
FEATURE      236
#active_site Lys (covalent AMP-binding) #status
            predicted
SUMMARY     #length 564 #molecular-weight 64472 #checksum 7075
Query Match 38.6%; Score 68; DB 2; Length 564;
Best Local Similarity 40.0%; Pred. No. 4.22e+00;
Matches 10; Conservative 8; Mismatches 6; Indels 1; Gaps 1;
Db 79 MYKVINIGDVAYVIGSLKKSIV 103
   |||::|||::|||::|||
Qy 1 MEKFMFAEFGQ-GYVOTPF 24

RESULT      4
ENTRY       Q0BE34 #type complete
TITLE       BBLF4 protein - human herpesvirus 4 (strain B95-8)
ORGANISM    #formal_name human herpesvirus 4, Epstein-Barr virus
DATE        25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change

```

```

ACCESSIONS F43043; A03776; S33026
REFERENCE   A93065
#authors    Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
            Mol. Biol. Med. (1983) 1:21-45
#title      Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8
            Epstein-Barr virus.
#cross-references MUID:85035713
#accession  F43043
#molecule_type DNA
#residues   1-809 #label BAN
#cross-references EMBL:V01555; NID:g59074; PID:g1334885
REFERENCE   A03794
#authors    Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.;
            Farrell, P.J.; Gibson, T.J.; Hatfull, G.; Hudson, G.S.;
            Satchwell, S.C.; Seguin, C.; Tuffnell, P.S.; Barrell, B.G.
            Nature (1984) 310:207-211
#journal    DNA sequence and expression of the B95-8 Epstein-Barr virus
            genome.
#title
#cross-references MUID:84270667
#contents   annotation; protein coding region
CLASSIFICATION #superfamily varicella-zoster virus gene 55 protein
SUMMARY     #length 809 #molecular-weight 89853 #checksum 6134
Query Match 38.6%; Score 68; DB 1; Length 809;
Best Local Similarity 42.18; Pred. No. 4.22e+00;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
Db 46 METFSPFDPFELSEPPFLP 64
   ||| |||
Qy 1 MEKFMFAEFGQGYVOTPF 19

RESULT      5
ENTRY       A25884 #type complete
TITLE       DNA-directed RNA polymerase (EC 2.7.7.6) II 140K chain -
            yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES DNA-directed RNA polymerase B chain B150; protein O3533;
            protein YOR151c
ORGANISM    #formal_name Saccharomyces cerevisiae
DATE        16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change
ACCESSIONS A25884; S36860; A37894; S67039
REFERENCE   A25884
#authors    Sweetser, D.; Nonet, M.; Young, R.A.
            Proc. Natl. Acad. Sci. U.S.A. (1987) 84:1192-1196
#journal    Prokaryotic and eukaryotic RNA polymerases have homologous
            core subunits.
#title
#cross-references MUID:87147239
#accession  A25884
#molecule_type DNA
#residues   1-1224 #label SWE
#cross-references EMBL:M15693
REFERENCE   S36860
#authors    Sweetser, D.; Nonet, M.; Young, R.A.
            Submitted to the EMBL Data Library, July 1987
#submission Prokaryotic and eukaryotic RNA polymerases have homologous
            core subunits.
#description
#accession  S36860
#molecule_type DNA
#residues   1-1002,'RRRY',1007-1224 #label SW2
#cross-references EMBL:M15693; NID:g172210; PID:g172211
REFERENCE   A37894
#authors    Riva, M.; Carles, C.; Sentenac, A.; Grachev, M.A.; Mustaev,
            A.A.; Zaychikov, E.F.
            J. Biol. Chem. (1990) 265:16498-16503
#journal    Mapping the active site of yeast RNA polymerase B (II).
#title
#cross-references MUID:90375520
#accession  A37894
#molecule_type protein
#residues   961-990 #label RIV
REFERENCE   S67032
#authors    Bordonne, R.; Camasses, A.; Madania, A.; Martin, R.P.; Poch,

```

```

#accession S35168
##molecule_type mRNA
##residues 1-56 ##label MEI
##cross-references EMBL:X69775; NID:g395301; PID:g395302
CLASSIFICATION ##superfamily unassigned cytochrome P450; cytochrome P450
homology
KEYWORDS chromoprotein; electron transfer; heme; iron; monooxygenase;
oxidoreductase
FEATURE
2
SUMMARY
#binding_site heme iron (Cys) (axial ligand) #status
predicted
#length 56 #checksum 2986
Query Match 37.5%; Score 66; DB 2; Length 56;
Best Local Similarity 23.4%; Pred. No. 8.42e+00;
Matches 5; Conservative 9; Mismatches 3; Indels 0; Gaps 0;
Db 20 LQREKFDVAPSYVHAPF 36
:::|: : : :|:|:|
Qy 1 MEKFMAEFGQGYVQTFF 17
RESULT 8
ENTRY E64757 #type complete
TITLE probable membrane protein b0309 - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
24-Oct-1998
ACCESSIONS E64757
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shoat
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession E64757
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-70 ##label BLAT
##cross-references GB:A5000137; GB:U00096; NID:g2367108; PID:g1786500;
UWGP:b0309
##experimental_source strain K-12, substrain MG1655
transmembrane protein
KEYWORDS
FEATURE
34-50 #domain transmembrane #status predicted #label TMM
SUMMARY #length 70 #molecular_weight 7854 #checksum 6854
Query Match 37.5%; Score 66; DB 2; Length 70;
Best Local Similarity 53.3%; Pred. No. 8.42e+00;
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
Db 47 QVYVQTSYLSQQSSI 61
| | | | : | | : | : |
Qy 10 QGYVQTPFLSESNV 24
RESULT 9
ENTRY GNBPT4 #type complete
TITLE gene 33 protein - phage T4
ORGANISM #formal_name phage T4
#note host Escherichia coli
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
05-Sep-1997
ACCESSIONS S05558
REFERENCE S05555
#authors Hahn, S.; Rueger, W.
#journal Nucleic Acids Res. (1989) 17:6729
#title Organization of the bacteriophage T4 genome between map
positions 150.745 and 145.824.
#cross-references MUID:89386003

```

```

Primary structure of the Pseudomonas enzyme.
#cross-references MUID:92241278
#accession S21209
##status preliminary
##molecule_type protein
##residues 1-357 ##label RUE
SUMMARY #length 357 #molecular-weight 40060 #checksum 6289

Query Match 37.5%; Score 66; DB 2; Length 357;
Best Local Similarity 38.1%; Pred. No. 8.42e+00;
Matches 8; Conservative 7; Mismatches 5; Indels 1; Gaps 1;

Db 226 IEFLMQFNGEGIQHVAFLSD 246
      :| | :| | :| | :| | :| | :| |
QY 1 MEKFMAEFGGVVQTPLSE 20

RESULT 12
ENTRY I64032 #type complete
TITLE hypothetical protein Hii500 - Haemophilus influenzae (strain Rd KW20)
ORGANISM #formal_name Haemophilus influenzae
DATE 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 30-Jun-1998
ACCESSIONS I64032
REFERENCE A64000
AUTHORS Fleisichmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kervatage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.; FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton, D.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek, M.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrman, J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C. Science (1995) 269:496-512
Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
#journal
#title
#cross-references MUID:95350630
#accession I64032
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-508 ##label TIGR
##cross-references GB:U32826; GB:L42023; NID:g1574322; PID:g1574341; TIGR:Hii500
SUMMARY #length 508 #molecular-weight 57190 #checksum 7619

Query Match 37.5%; Score 66; DB 2; Length 508;
Best Local Similarity 43.8%; Pred. No. 8.42e+00;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 411 MPKYKALYESGYTOIP 426
      | | :| | :| | :| |
QY 1 MEKFMAEFGGVVQTP 16

RESULT 13
ENTRY H64775 #type complete
TITLE probable membrane protein ylab - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 18-Sep-1998
ACCESSIONS H64775
REFERENCE A64720
AUTHORS Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shoemaker, Y. Science (1997) 277:1453-1462
The complete genome sequence of Escherichia coli K-12.
#journal
#title
#cross-references MUID:97426617

#cross-references translation not shown
#accession S05558
##status preliminary
##molecule_type DNA
##residues 1-112 ##label HAH
##cross-references EMBL:X15818; NID:g15210; PID:g15213
GENETICS
#gene 33
CLASSIFICATION #superfamily phage T4 gene 33 protein
SUMMARY #length 112 #molecular-weight 12830 #checksum 9958

Query Match 37.5%; Score 66; DB 1; Length 112;
Best Local Similarity 46.2%; Pred. No. 8.42e+00;
Matches 12; Conservative 7; Mismatches 4; Indels 3; Gaps 2;

Db 44 IEKMVEFGMSYLEATTAFLSE-NSI 68
      :| | :| | :| | :| | :| | :| |
QY 1 MEKFMAEFGGVVQ--TPFLSESNSV 24

RESULT 10
ENTRY B71121 #type complete
TITLE hypothetical protein PH0737 - Pyrococcus horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
ACCESSIONS B71121
REFERENCE A71000
AUTHORS Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuoka, H.; Kikuchi, H. DNA Res. (1998) 5:55-76
Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.
#journal
#title
#cross-references MUID:98344137
#accession B71121
##status preliminary; nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-336 ##label KAW
##cross-references GB:AP00003; NID:g3236130; PID:d1030771; PID:g3257145
##experimental_source strain OT3
##note this accession replaces an interim accession for a sequence replaced by GenBank
GENETICS
#gene PH0737
SUMMARY #length 336 #molecular-weight 37579 #checksum 7869

Query Match 37.5%; Score 66; DB 2; Length 336;
Best Local Similarity 38.9%; Pred. No. 8.42e+00;
Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Db 166 EYNGFVKSHFLDDKASV 183
      | | :| | :| | :| |
QY 7 EFQGGYVQTPFLSESNSV 24

RESULT 11
ENTRY S21209 #type complete
TITLE 4-hydroxyphenylpyruvate dioxygenase - Pseudomonas sp.
ORGANISM #formal_name Pseudomonas sp.
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
ACCESSIONS S21209
REFERENCE Ruetschi, U.; Odelhoeg, B.; Lindstedt, S.; Bartos-Sooderling, J.; Persson, B.; Joernvall, H. Eur. J. Biochem. (1992) 205:459-466
Characterization of 4-hydroxyphenylpyruvate dioxygenase.
#journal
#title

```

#accession H64775  
#status nucleic acid sequence not shown; translation not shown  
#molecule\_type DNA  
#residues 1-518 #label BLAT  
#cross-references GB:AE000152; GB:U00096; NID:g1786660; PID:g1786662;  
#experimental\_source strain K-12, substrain MG1655  
UNGP:b0457

## GENETICS

#gene ylaB  
CLASSIFICATION #superfamily probable membrane protein ylaB  
KEYWORDS transmembrane protein  
FEATURE 8-24 #domain transmembrane #status predicted #label TM1\  
251-267 #domain transmembrane #status predicted #label TM2  
SUMMARY #length 518 #molecular-weight 58932 #checksum 2466

Query Match 37.5%; Score 66; DB 2; Length 518;  
Best Local Similarity 31.6%; Pred. No. 8.42e+00;  
Matches 6; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

Db 420 EYLDDEFTGYSSLSYLQD 438  
| :: :||| :: :|| :  
QY 2 EKFMAEFGQGYVQTPLSE 20

RESULT 14  
ENTRY H70470 #type complete  
TITLE hypothetical protein aq\_1989 - Aquifex aeolicus  
ORGANISM #formal\_name Aquifex aeolicus  
DATE 08-May-1998 #sequence\_revision 08-May-1998 #text\_change  
08-May-1998

ACCESSIONS H70470  
REFERENCE A70300  
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;  
Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;  
Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,  
J.M.; Olson, G.J.; Swanson, R.V.  
#journal Nature (1998) 392:353-358  
#title The complete genome of the hyperthermophilic bacterium  
Aquifex aeolicus.  
#cross-references MUID:98196666

#accession H70470  
#status preliminary; nucleic acid sequence not shown;  
translation not shown  
#molecule\_type DNA  
#residues 1-853 #label AQF  
#cross-references GB:AE000766; NID:g2984216; PID:g2984229; GB:AE000657  
#experimental\_source strain VF5

## GENETICS

#gene aq\_1989  
SUMMARY #length 853 #molecular-weight 99499 #checksum 642  
Query Match 37.5%; Score 66; DB 2; Length 853;  
Best Local Similarity 35.0%; Pred. No. 8.42e+00;  
Matches 7; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

Db 586 LEIKVKDFEEKYPSPLPE 605  
| :: :||| :|| :|| :  
QY 1 MEKFMAEFGQGYVQTPLSE 20

RESULT 15  
ENTRY S65068 #type complete  
TITLE DNA-directed RNA polymerase (EC 2.7.7.6) II second largest  
chain - tomato  
ORGANISM #formal\_name Lycopersicon esculentum #common\_name tomato  
DATE 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change  
21-Aug-1998

ACCESSIONS S65068  
REFERENCE S65068  
#authors Warrilow, D.; Symons, R.H.  
#journal Plant Mol. Biol. (1996) 30:337-342  
#title Sequence analysis of the second largest subunit of tomato RNA

polymerase II.  
#cross-references MUID:96178872  
#accession S65068  
#status nucleic acid sequence not shown  
#molecule\_type mRNA  
#residues 1-1191 #label WAR  
#cross-references EMBL:U28403; NID:g1049067; PID:g1049068

## GENETICS

#gene RPB2  
CLASSIFICATION #superfamily DNA-directed RNA polymerase 132K polypeptide  
KEYWORDS DNA binding; nucleotidyltransferase; nucleus; transcription  
SUMMARY #length 1191 #molecular-weight 135063 #checksum 5669

Query Match 37.5%; Score 66; DB 2; Length 1191;  
Best Local Similarity 35.0%; Pred. No. 8.42e+00;  
Matches 7; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

Db 85 YKINFGQIYLSKPMMTESDG 104  
| :: :||| | :|| :  
QY 4 FMAEFGQGYVQTPLSESNS 23

Search completed: Wed Sep 1 16:06:24 1999.  
Job time : 14 secs.

\*\*\*\*\*  
W E S L E H  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Sep 1 16:05:00 1999; MasPar time 2.92 Seconds  
Tabular output not generated. 232.136 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (1-24) from PCTUS9913024.pep (2 of 12)  
Perfect Score: 176  
Sequence: 1 MEKFMAGGQYVQTPFLSESNSV 24

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 31.468; Variance 47.304; scale 0.665

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	74	42.0	65	1 Y15A_BPT4	HYPOTHETICAL 7.3 KD PR	1.45e+01
2	70	39.8	923	1 RE11_SCHPO	MEIOTIC RECOMBINATION	7.11e+01
3	68	38.6	564	1 DNLI_FOWPM	DNA LIGASE (EC 6.5.1.1)	1.54e+00
4	68	38.6	809	1 HELI_EBV	PROBABLE HELICASE.	1.54e+00
5	68	38.6	1224	1 RP22_YEAST	DNA-DIRECTED RNA POLYM	1.54e+00
6	67	38.1	236	1 COAT_MCMV	COAT PROTEIN (CAPSID P	2.25e+00
7	66	37.5	112	1 VG33_BPT4	RNA POLYMERASE-ASSOCIA	3.28e+00
8	66	37.5	357	1 HPD2_PSESP	4-HYDROXYPHENYLPYRUVAT	3.28e+00
9	66	37.5	508	1 YF00_HAEIN	HYPOTHETICAL PROTEIN H	3.28e+00
10	66	37.5	518	1 YLAB_ECOLI	HYPOTHETICAL 58.9 KD P	3.28e+00
11	66	37.5	524	1 CP72_CATRO	CYTOCROME P450 72A1 (	3.28e+00
12	65	36.9	277	1 MALD_STRPN	MALTODEXTRIN TRANSPORT	4.76e+00
13	64	36.4	239	1 YQCB_HAEIN	HYPOTHETICAL PROTEIN H	6.88e+00
14	64	36.4	420	1 PGKE_TRYBB	PHOSPHOGLYCERATE KINAS	6.88e+00
15	64	36.4	440	1 PGKC_TRYBB	PHOSPHOGLYCERATE KINAS	6.88e+00
16	64	36.4	505	1 PGKA_TRYBB	PHOSPHOGLYCERATE KINAS	6.88e+00
17	64	36.4	508	1 PGKD_TRYBB	PHOSPHOGLYCERATE KINAS	6.88e+00
18	64	36.4	556	1 INET_YEAST	HYPOTHETICAL 74.8 KD P	6.88e+00
19	64	36.4	759	1 TRP2_SCHPO	ANTHRANILATE SYNTHASE	6.88e+00
20	64	36.4	1188	1 RP22_ARATH	DNA-DIRECTED RNA POLYM	6.88e+00
21	63	35.8	389	1 KSS5_ECOLI	CAPSULE POLYSACCHARIDE	9.89e+00
22	63	35.8	389	1 ALR_BACSU	ALANINE RACEMASE (EC 5	9.89e+00
23	63	35.8	417	1 PGKB_CRIFA	PHOSPHOGLYCERATE KINAS	9.89e+00

24 63 35.8 455 1 PGKC\_CRIFA PHOSPHOGLYCERATE KINAS 9.89e+00  
25 63 35.8 603 1 VEL\_HPVB8 REPLICATION PROTEIN E1 9.89e+00  
26 63 35.8 686 1 YIK5\_YEAST FUMARATE REDUCTASE 13 9.89e+00  
27 62 35.2 119 1 FRDD\_PROVU FUMARATE REDUCTASE 13 1.42e+01  
28 62 35.2 402 1 OPDE\_PSEAE TRANSCRIPTION REGULATO 1.42e+01  
29 62 35.2 417 1 PGKB\_LEIMA PHOSPHOGLYCERATE KINAS 1.42e+01  
30 62 35.2 421 1 PGKB\_TRYBB PHOSPHOGLYCERATE KINAS 1.42e+01  
31 62 35.2 479 1 PGKC\_LEIMA PHOSPHOGLYCERATE KINAS 1.42e+01  
32 62 35.2 555 1 HYFG\_ECOLI HYDROGENASE-4 COMPONENT 1.42e+01  
33 62 35.2 569 1 HYCE\_ECOLI FORMATE HYDROGENLYASE 1.42e+01  
34 62 35.2 604 1 VEL\_HPVB36 REPLICATION PROTEIN E1 1.42e+01  
35 62 35.2 652 1 SSPI\_SCHPO SERINE/THREONINE-PROTE 1.42e+01  
36 62 35.2 858 1 UBPS\_HUMAN UBIQUITIN CARBOXYL-TER 1.42e+01  
37 62 35.2 858 1 UBPS\_MOUSE UBIQUITIN CARBOXYL-TER 1.42e+01  
38 61 34.7 96 1 Y112\_SSV1 HYPOTHETICAL 11.2 KD P 2.02e+01  
39 61 34.7 297 1 KPRS\_MYCGE RIBOSE-PHOSPHATE PYROP 2.02e+01  
40 61 34.7 476 1 VTDB\_RABIT VITAMIN D-BINDING PROT 2.02e+01  
41 61 34.7 610 1 CHIT\_STRPL CHITINASE 63 PRECURSOR 2.02e+01  
42 61 34.7 659 1 HEP4\_HSV7J DNA HELICASE/PRIMASE C 2.02e+01  
43 61 34.7 676 1 VI08\_VACCV PUTATIVE RNA HELICASE 2.02e+01  
44 61 34.7 676 1 VI08\_VARV PUTATIVE RNA HELICASE 2.02e+01  
45 61 34.7 762 1 AMO2\_BOVIN COPPER AMINE OXIDASE P 2.02e+01

## ALIGNMENTS

RESULT 1  
ID Y15A\_BPT4 STANDARD; PRT; 65 AA.  
AC P39509;  
DT 01-FEB-1995 (REL. 31, CREATED)  
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DE HYPOTHETICAL 7.3 KD PROTEIN IN GP34-GP35 INTERGENIC REGION.  
GN Y15A OR 34.1.  
OS BACTERIOPHAGE T4.  
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE;  
OC T4-LIKE PHAGES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KUTTER E.M.;  
RL SUBMITTED (NOV-1994) TO THE SWISS-PROT DATA BANK.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 65 AA; 7334 MW; C2D7CE2D CRC32;

Query Match 42.0%; Score 74; DB 1; Length 65;  
Best Local Similarity 58.3%; Pred. No. 1.45e+01;  
Matches 14; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Db 1 MEKFMAGGQYVQTPFLSESNSV 24  
QY 1 MEKFMAGGQYVQTPFLSESNSV 24  
||||||| :|||||||

RESULT 2  
ID RE11\_SCHPO STANDARD; PRT; 923 AA.  
AC Q92380;

DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DE MEIOTIC RECOMBINATION PROTEIN RECI1.  
GN RECI1.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;  
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;  
OC SCHIZOSACCHAROMYCES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE; 97231330.  
RX LI Y.F., NUMATA M., WAHLS W.P., SMITH G.R.;

\*Region-specific meiotic recombination in Schizosaccharomyces pombe:  
the rec11 gene.;  
RL MOL. MICROBIOL. 23:869-878(1997).  
-----

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U70737; G1619901; -
KW MEIOSIS.
SQ SEQUENCE 923 AA; 107418 MW; B51C7725 CRC32;

Query Match 39.8%; Score 70; DB 1; Length 923;
Best Local Similarity 42.1%; Pred. No. 7.11e+01;
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

Db 186 FLVEFISKVYVEDAFGLGS 204
Qy 4 FMAEF-GOGYVOTPFLES 21

RESULT 3
ID DNL1.FOWPM STANDARD; PRT; 564 AA.
AC 067480;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DNA LIGASE (EC 6.5.1.1) (POLYDEOXYRIBONUCLEOTIDE SYNTHASE (ATP)).
GN ASOR.
OS FOWLPX VIRUS (ISOLATE HP-438[MUNICH]).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; FOXVIRIDAE; CHORDOPOXVIRINAE;
OC AVIPOXVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94358756.
RA SKINNER M.A., MOORE J.B., BINNS M.M., SMITH G.L., BOURSNEILL M.E.;
RT "Deletion of fowlpox virus homologues of vaccinia virus genes between
RT the 3 beta-hydroxysteroid dehydrogenase (A44L) and DNA ligase (A50R)
RT genes.";
RL J. GEN. VIROL. 75:2495-2498(1994).
CC -!- FUNCTION: THIS PROTEIN SEALS, DURING DNA REPLICATION, DNA
CC RECOMBINATION AND DNA REPAIR, NICKS IN DOUBLE-STRANDED DNA.
CC IT IS NOT ESSENTIAL FOR VIRAL REPLICATION AND RECOMBINATION.
CC -!- CATALYTIC ACTIVITY: ATP + (DEOXYRIBONUCLEOTIDE)(N) +
CC (DEOXYRIBONUCLEOTIDE)(M) -> AMP + PYROPHOSPHATE +
CC (DEOXYRIBONUCLEOTIDE)(N+M).
CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT DNA LIGASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 229716; G453602; -
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
DR PROSITE; PS00333; DNA_LIGASE_A2; FALSE_NEG.
DR PFAM; PF01058; DNA_Ligase; 1.
KW DNA REPAIR; DNA REPLICATION; DNA RECOMBINATION; CELL DIVISION; LIGASE;
ATP-BINDING.
FT BINDING. 225 225 AMP (BY SIMILARITY).
SQ SEQUENCE 564 AA; 64472 MW; 861BEDB9 CRC32;

Query Match 38.6%; Score 68; DB 1; Length 564;
Best Local Similarity 40.0%; Pred. No. 1.54e+00;
Matches 10; Conservative 8; Mismatches 6; Indels 1; Gaps 1;

Db 79 MYKYVINIGDVAVYIGSFLKSKSV 103
Qy 1 MEKFMAEFGG-GYVQTFFLESNSV 24
```

```
RESULT 4
ID HELI.EBV STANDARD; PRT; 809 AA.
AC P03214;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
DE PROBABLE HELICASE.
GN BBLF4.
OS EPSTEIN-BARR VIRUS (STRAIN B95-8) (HUMAN HERPESVIRUS 4).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE; LYMPHOCRYPTOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84270667.
RA BAER R., BANKIER A.T., BIGGIN M.D., DEININGER P.L., FARRELL P.J.,
RA GIBSON T.J., HATFULL G., HUDSON G.S., SATCHWELL S.C., SEGUIN C.,
RA TUFFNELL P.S., BARRELL B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL NATURE 310:207-211(1984).
CC -!- FUNCTION: THIS PROTEIN MAY BE AN HELICASE AND IS REQUIRED FOR
CC REPLICATION OF VIRAL DNA.
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL5,
CC EHV-1 57, EBV BBLF4, HCMV UL105, AND VZV 55.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V01555; G59126; -
DR PIR; A03776; Q0BE34.
DR PIR; S33026; S33026.
KW DNA REPLICATION; ATP-BINDING; HELICASE; EARLY PROTEIN.
FT NP-BIND 72 79 ATP (BY SIMILARITY).
SQ SEQUENCE 809 AA; 89853 MW; F70F3487 CRC32;

Query Match 38.6%; Score 68; DB 1; Length 809;
Best Local Similarity 42.1%; Pred. No. 1.54e+00;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Db 46 METFSPEFDPSELSEPPFLP 64
Qy 1 MEKFMAEFGG-GYVQTFFLS 19

RESULT 5
ID RPB2.YEAST STANDARD; PRT; 1224 AA.
AC P08518; Q12738;
DT 01-AUG-1988 (REL. 08, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DNA-DIRECTED RNA POLYMERASE II 140 KD POLYPEPTIDE (EC 2.7.7.6) (B150)
DE (RNA POLYMERASE II SUBUNIT 2).
GN RPB2 OR RPO22 OR RPB150 OR YOR151C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87147239.
RA SWEETSER D., NONET M., YOUNG R.A.;
RT "Prokaryotic and eukaryotic RNA polymerases have homologous core
RT subunits.";
RL PROC. NATL. ACAD. SCI. U.S.A. 84:1192-1196(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1678;
RA AYADI A., BORDONNE R., CANASSES A., MADANIA A., POCH O.,
RA TARASSOV I.A., WINSOR B., MARTIN R.P.;
```

RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
CC SUBSTRATES.  
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +  
CC RNA(N).  
CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
CC THIS SUBUNIT IS THE SECOND LARGEST COMPONENT OF RNA POLYMERASE II.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE FOUND IN  
CC EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA PRECURSOR,  
CC POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE III FOR 5S  
CC AND TRNA GENES.  
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M15693; G172211; -  
CC DR EMBL: U55020; G1293711; -  
CC DR EMBL: Z75059; E252039; -  
CC DR PIR: A25884; A25884.  
CC DR SGD: L0001676; RPB2.  
CC DR PROSITE: PS01166; RNA\_POL\_BETA; 1.  
CC DR PFAM: PF00562; RNA\_POL\_B; 1.  
KW DNA-DIRECTED RNA POLYMERASE; TRANSCRIPTION; ZINC; ZINC-FINGER;  
KW NUCLEAR PROTEIN.  
FT ZN\_FING 1163 1185 C4-TYPE (POTENTIAL).  
FT CONFLICT 1003 1006 AEG1 -> RRRY (IN REF. 1).  
SQ SEQUENCE 1224 AA; 138751 MW; 138751 MW; DA2ED93A CRC32;  
  
Query Match 38.68; Score 68; DB 1; Length 1224;  
Best Local Similarity 40.94; Pred. No. 1.54e+00;  
Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;  
  
Db 87 KYEISFGKIVYTKPMVNESDGV 108  
QY 3 KFWAEFGQGVQTPFLSENSV 24  
  
RESULT 6  
ID COAT\_MCMV STANDARD; PRT; 236 AA.  
AC P11642;  
DT 01-OCT-1989 (REL. 12, CREATED)  
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)  
DE COAT PROTEIN (CAPSID PROTEIN).  
OS MAIZE CHLOROTIC MOTTLE VIRUS (MCMV).  
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; MACHLOMOVIRUS.  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE: 89263725.  
RA NUTTER R.C., SCHEETS K., PANGANIBAN L.C., LOMMEL S.A.;  
RT "The complete nucleotide sequence of the maize chlorotic mottle virus  
RT genome".  
CC  
CC NUCLEIC ACIDS RES. 17:3163-3177(1989).  
CC -1- SIMILARITY: COAT PROTEINS FROM CARMOVIRUSES (CARMV, MNSV, TCV),  
CC DIANTHOVIRUSES (RCNMV), SOBEMOVIRUSES (MCMV, SBMV), TOMBUSVIRUSES  
CC (AMCV, CNV, CYRSV, AND TBSV), AND TNV ARE RELATED IN THEIR S  
CC REGION.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

CC EMBL: X14736; G60533; -  
CC DR PIR: JQ0061; JQ0061.  
CC DR PROSITE: PS00555; ICOSAH\_VIR\_COAT\_S; 1.  
CC DR PFAM: PF00729; Viral\_coat; 1.  
KW COAT PROTEIN. 1 51 R DOMAIN, INTERACTION WITH RNA.  
FT DOMAIN 52 217 S DOMAIN, VIRION SHELL.  
FT DOMAIN 218 236 P DOMAIN, PROJECTING.  
SQ SEQUENCE 236 AA; 25149 MW; 647462A9 CRC32;  
  
Query Match 38.18; Score 67; DB 1; Length 236;  
Best Local Similarity 42.18; Pred. No. 2.29e+00;  
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
  
Db 111 EFTADTTSGYISMAFLSD 129  
QY 2 EKFWAEFGQGVQTPFLSE 20  
  
RESULT 7  
ID VG33\_BPT4 STANDARD; PRT; 112 AA.  
AC P13338;  
DT 01-JAN-1990 (REL. 13, CREATED)  
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
DE RNA POLYMERASE-ASSOCIATED PROTEIN GP33.  
GN 33.  
OS BACTERIOPHAGE T4.  
OC VIRUSES; DSNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE;  
OC T4-LIKE PHAGES.  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-BK536;  
RX MEDLINE: 89386003.  
RA HAHN S., RUEGER W.;  
RT "Organization of the bacteriophage T4 genome between map positions  
RT 150.745 and 145.824".  
RL NUCLEIC ACIDS RES. 17:6729-6729(1989).  
RN [2]  
RN CHARACTERIZATION.  
RX MEDLINE: 94148776.  
RA WINKELMAN J.W., KASSAVETIS G.A., GEIDUSCHEK E.P.;  
RT "Molecular genetic analysis of a prokaryotic transcriptional  
RT coactivator: functional domains of the bacteriophage T4 gene 33  
RT protein".  
RL J. BACTERIOL. 176:1164-1171(1994).  
CC -1- FUNCTION: BINDS TO E.COLI RNA POLYMERASE AND IS REQUIRED FOR  
CC TRANSCRIPTION OF LATE GENES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X15818; G15213; -  
CC DR PIR: S05558; GWP74.  
SQ SEQUENCE 112 AA; 12830 MW; 5B4A3D25 CRC32;  
  
Query Match 37.58; Score 66; DB 1; Length 112;  
Best Local Similarity 46.28; Pred. No. 3.28e+00;  
Matches 12; Conservative 7; Mismatches 4; Indels 3; Gaps 2;  
  
Db 44 IERNVEEFGMSYLEATTAFLSE-NSI 68  
QY 1 MEKFWAEFGQGVQ--TPFLSENSV 24  
  
RESULT 8  
ID HPPD\_PSESP STANDARD; PRT; 357 AA.

```

SQ SEQUENCE 508 AA; 57190 MW; 98DDC6AA CRC32;

Query Match 37.5%; Score 66; DB 1; Length 508;
Best Local Similarity 43.8%; Pred. No. 3.28e+00;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 411 MPYKALYESGYIOIP 426
QY 1 MEKFAEFGGQVOTP 16
I : I : I : I : I : I :
| : | : | : | : | : |

RESULT 10
ID YLAB_ECOLI STANDARD; PRT; 518 AA.
AC P77473;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 58.9 KD PROTEIN IN TESB-HHA INTERGENIC REGION.
YLAB.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERiaceae;
OC ESCHERICHIA.
[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL SCIENCE 277:1453-1474(1997).
[2]
RN RP SEQUENCE FROM N.A.
RA ROBERTS D., ALLEN E., ARAUJO R., APARICIO A., CHUNG E., DAVIS K.,
RA DUNCAN M., FEDERSPEL N., HYMAN R., KALMAN S., KOMP C., KURDI O.,
RA LEW H., LIN D., NAMATH A., OEFFNER P., SCHRAMM S., DAVIS R.W.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: BELONGS TO THE YEGE/YHDA/YHJK/YJCC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AE000152; G1786662; -.
DR DR EMBL; U82664; G1773140; -.
DR ECGENE; EG14237; YLAB.
DR PFAM; PF00563; DUF2.1.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 8 28
FT TRANSNUM 244 264 POTENTIAL.
FT SEQUENCE 518 AA; 58932 MW; 80D18559 CRC32;

Query Match 37.5%; Score 66; DB 1; Length 518;
Best Local Similarity 31.6%; Pred. No. 3.28e+00;
Matches 6; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

Db 420 EYLDDFGTGYSSLSYLQD 438
QY 2 EKFAEFGGQVOTPFLE 20
I : I : I : I : I : I :
| : | : | : | : | : |

RESULT 11
ID CP72_CATRO STANDARD; PRT; 524 AA.
AC Q05047;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450_7A21;

```







DR PDB; 13PK; 24-DEC-97.  
DR PDB; 16PK; 25-NOV-98.  
DR PROSITE; PS00111; GLYCERATE\_KINASE; 1.  
DR PFAM; PF00162; PKG; 1.  
KW TRANSFERASE: KINASE; GLYCOLYSIS; MULTIGENE FAMILY; GLYCOSOME;  
3D-STRUCTURE.  
FT VARIANT 75 75 G -> D (IN ALLELE 4).  
SQ SEQUENCE 440 AA; 47118 MW; 4A1F7393 CRC32;

Query Match 36.4%; Score 64; DB 1; Length 440;  
Best Local Similarity 35.3%; Pred. No. 6.88e+00;  
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 151 MAKILASYGDVYISDAF 167  
| | : | : | : |  
QY 1 MEKFMAEFGGYVQTPF 17

Search completed: Wed Sep 1 16:05:11 1999  
Job time : 11 secs.

\*\*\*\*\*  
 W P S R F A  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Blocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Wed Sep 1 16:05:29 1999; MasPar time 5.87 Seconds  
 Tabular output not generated. 223.019 Million cell updates/sec

Title: >PCT-US99-13024-2  
 Description: (1-24) from PCTUS9913024.pep (2 of 12)  
 Perfect Score: 176  
 Sequence: 1 MEKMAEFQGGYVQTPFLSESNSV 24

Scoring table: PAM 150  
 Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: sptrembl9

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
 5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
 9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
 13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 29.931; Variance 49.849; scale 0.600

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	77	43.8	860	1	OS9003 860AA LONG HYPOTHETICAL	1.28e-01
2	75	42.6	364	3	074624 MANNOSYL-1-PHOSPHATE GU	2.75e-01
3	74	42.0	178	3	050074 HYPOTHETICAL 19.4 KD P	4.03e-01
4	73	41.5	2301	10	080784 FL13P17.19 PROTEIN.	5.87e-01
5	69	39.2	1132	4	P78418 KIAA0287 (PEG3) (FRAGM	2.58e+00
6	68	38.6	418	2	056631 LECITHINASE.	3.70e+00
7	68	38.6	470	2	087325 LECITHINASE.	3.70e+00
8	68	38.6	1378	11	Q51138 PATERNALLY EXPRESSED P	3.70e+00
9	68	38.6	1571	11	Q34978 ZINC FINGER PROTEIN.	3.70e+00
10	66	37.5	70	2	P75688 FROM BASES 311709 TO 3	7.55e+00
11	66	37.5	215	2	087765 PYRROLIDONE CARBOXYL P	7.55e+00
12	66	37.5	296	2	Q47402 REGION 1 GENE PRODUCT	7.55e+00
13	66	37.5	334	5	Q23414 COSMID ZK1248.	7.55e+00
14	66	37.5	336	1	038468 336AA LONG HYPOTHETICAL	7.55e+00
15	66	37.5	516	10	Q42701 CYTOCHROME P450 (EC 1.	7.55e+00
16	66	37.5	853	2	067796 HYPOTHETICAL 99.5 KD P	7.55e+00
17	66	37.5	1191	10	Q42877 RNA POLYMERASE II SUBU	7.55e+00
18	66	37.5	2219	5	Q23398 ZK1067.2 PROTEIN.	7.55e+00
19	65	36.9	423	2	Q24752 RIBA PROTEIN.	1.07e-01
20	65	36.9	526	2	Q60102 PERIPLASMIC ALPHA-AMYL	1.07e-01

21 65 36.9 591 2 069736 HYPOTHETICAL 64.6 KD P 1.07e+01  
 22 64 36.4 273 2 067454 HYPOTHETICAL 32.9 KD P 1.52e+01  
 23 64 36.4 355 5 023003 SIMILAR TO FAMILY 1 OF 1.52e+01  
 24 64 36.4 357 2 048225 MAJOR OUTER MEMBRANE P 1.52e+01  
 25 64 36.4 420 8 037744 PHOSPHOGLYCERATE KINAS 1.52e+01  
 26 64 36.4 440 8 037745 PHOSPHOGLYCERATE KINAS 1.52e+01  
 27 64 36.4 509 8 037743 PHOSPHOGLYCERATE KINAS 1.52e+01  
 28 64 36.4 524 10 042700 CYTOCHROME P450 (EC 1. 1.52e+01  
 29 64 36.4 597 2 033086 HYPOTHETICAL 65.2 KD P 1.52e+01  
 30 64 36.4 980 2 024780 CELLODEXTRIN PHOSPHORY 1.52e+01  
 31 63 35.8 118 14 065377 DNA FOR EARLY REGION 1 2.14e+01  
 32 63 35.8 175 1 038668 HYPOTHETICAL 30.9 KD P 2.14e+01  
 33 63 35.8 278 2 006991 175AA LONG HYPOTHETICA 2.14e+01  
 34 63 35.8 455 2 084013 ACYLTRANSFERASE. 2.14e+01  
 35 63 35.8 645 2 053818 PPE-FAMILY PROTEIN. 2.14e+01  
 36 63 35.8 730 3 039248 CHROMOSOME XV READING 2.14e+01  
 37 63 35.8 815 2 047732 ORF16. 2.14e+01  
 38 63 35.8 1026 3 042948 HYPOTHETICAL 115.5 KD 2.14e+01  
 39 63 35.8 1215 2 048232 DNA FOR SEROTYPE B CAP 2.14e+01  
 40 63 35.8 1345 1 054437 STABLE PROTEASE PRECUR 2.14e+01  
 41 62 35.2 191 1 027457 FUCULOSE-1-PHOSPHATE A 3.01e+01  
 42 62 35.2 244 2 051694 MEMBRANE SPANNING PROT 3.01e+01  
 43 62 35.2 1165 2 045887 BOTULINUM NEUROTOXIN T 3.01e+01  
 44 62 35.2 1165 2 045844 NEUROTOXIN COMPLEX M M 3.01e+01  
 45 61 34.7 676 14 057193 NPH-II, HELICASE. 4.22e+01

#### ALIGNMENTS

RESULT 1  
 ID OS9003 PRELIMINARY; PRT; 860 AA.  
 AC OS9003:  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)  
 DE 860AA LONG HYPOTHETICAL PROTEIN.  
 GN PH1302.  
 OS PYROCOCUS HORIKOSHII.  
 OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OT3;  
 RX MEDLINE; 98344137.  
 RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,  
 RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOIYAMA A., NAGAI Y.,  
 RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAIYAMA M., OHFUKU Y.,  
 RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,  
 RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,  
 RA KIKUCHI H.;  
 RT \*Complete Sequence and Gene Organization of the Genome of a  
 RT Hyper-thermophilic Archaeobacterium, Pyrococcus horikoshii OT3.\*;  
 RL DNA RES. 5:55-76(1998).  
 DR EMBL; AP000006; D1031349; -  
 SQ SEQUENCE 860 AA; 95455 MW; B336F868 CRC32;

Query Match 43.8%; Score 77; DB 1; Length 860;  
 Best Local Similarity 29.4%; Pred. No. 1.28e-01;  
 Matches 5; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

Db 1 MSKELTDFSNFIRVAF 17  
 | | | | | : : : : :  
 QY 1 MEKMAEFQGGYVQTPF 17  
 | | | | | : : : : :  
 RESULT 2  
 ID 074624 PRELIMINARY; PRT; 364 AA.  
 AC 074624:  
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE MANNOSYL-1-PHOSPHATE GUANYLYLTRANSFERASE (EC 2.7.7.13).  
 GN MPGL.  
 OS HYPOCREA JECORINA.

[illegible]

Query Match 38.6%; Score 68; DB 2; Length 418;  
Best Local Similarity 33.3%; Pred. No. 3.70e+00;  
Matches 7; Conservative 9; Mismatches 5; Indels

Db 177 FLGHFNGFVMTYLAQGLNV 197  
 I:: I:: I:: I:: I:: I::  
 QY 4 FMAEFGGVQTPFLSESNSV 24

RESULT 7  
 ID O87325 PRELIMINARY; PRT: 470 AA.  
 AC O87325;  
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE LECITHINASE.  
 GN PHL.  
 OS VIBRIO MIMICUS.  
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; VIBRIACEAE; VIBRIO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 33653;  
 RX MEDLINE: 98440110.  
 RA KONG I.S.;  
 RT "Cloning and identification of a phospholipase gene from *Vibrio*  
 RT *mimicus*."  
 RL BIOCHIM. BIOPHYS. ACTA 1394:85-89(1998).  
 DR EMBL: AF035162; G3746409; -.  
 DR PROSITE: PS01098; LIPASE\_GDSL\_SER; 1.  
 SQ SEQUENCE 470 AA; 53194 MW; 1EF1ESA5 CRC32;

Query Match 38.6%; Score 68; DB 2; Length 470;

Best Local Similarity 33.3%; Pred. No. 3.70e+00; Mismatches 5; Indels 0; Gaps 0;

Db 229 FLGHFNGFVMTYLAQGLNV 249  
 I:: I:: I:: I:: I:: I::  
 QY 4 FMAEFGGVQTPFLSESNSV 24

RESULT 8  
 ID Q61138 PRELIMINARY; PRT: 1378 AA.  
 AC Q61138;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE PATERNALLY EXPRESSED PROTEIN 3 (2N-FINGER PROTEIN PW1).  
 GN PEG3.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SV129; TISSUE-LIMB BUD;  
 RX MEDLINE: 96400442.  
 RA RELAIX F., WENG X., MARAZZI G., YANG E., COPELAND N., JENKINS N.,  
 RA SPENCE S.E., SASSOON D.;  
 RT "Pw1, a novel zinc finger gene implicated in the myogenic and  
 RT neuronal lineages."  
 RL DEV. BIOL. 177:383-396(1996).  
 DR EMBL: G48804; G1197857; -.  
 DR MGD: MGI:104748; PEG3.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2; 11.  
 DR PFAM: PF00096; zf-C2H2; 11.  
 KW ZINC-FINGER; METAL-BINDING; DNA-BINDING.  
 SQ SEQUENCE 1378 AA; 156916 MW; 756F3644 CRC32;

Query Match 38.6%; Score 68; DB 11; Length 1378;

Best Local Similarity 40.0%; Pred. No. 3.70e+00; Mismatches 2; Indels 0; Gaps 0;

Db 1102 EYGPSYTHASFLTEP 1116  
 I:: I:: I:: I:: I:: I::  
 QY 7 EFGQGVQTPFLSES 21

RESULT 9  
 ID O54978 PRELIMINARY; PRT: 1571 AA.  
 AC O54978;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE ZINC FINGER PROTEIN.  
 GN PEG3.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 96154192.  
 RA KUROIWA Y., KANEKO-ISHINO T., KAGITANI F., KOHDA T., LI L.L., TADA M.,  
 RA SUZUKI R., YOKOYAMA M., SHIROISHI T., WAKANA S., BARTON S.C.,  
 RA ISHINO F., SURANI M.A.;  
 RT "Peg3 imprinted gene on proximal chromosome 7 encodes for a zinc  
 RT finger protein."  
 RL NAT. GENET. 12:186-190(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA SURANI A.M.;  
 RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AF038939; G2791678; -.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2; 11.  
 KW ZINC-FINGER; METAL-BINDING; DNA-BINDING.  
 SQ SEQUENCE 1571 AA; 178824 MW; 6F2820E6 CRC32;

Query Match 38.6%; Score 68; DB 11; Length 1571;  
 Best Local Similarity 40.0%; Pred. No. 3.70e+00; Mismatches 2; Indels 0; Gaps 0;

Db 1295 EYGPSYTHASFLTEP 1309  
 I:: I:: I:: I:: I:: I::  
 QY 7 EFGQGVQTPFLSES 21

RESULT 10  
 ID P75688 PRELIMINARY; PRT: 70 AA.  
 AC P75688;  
 DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
 DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE FROM BASES 311709 TO 323910  
 DE (SECTION 27 OF 400) OF THE COMPLETE GENOME (SECTION 27 OF 400).  
 OS ESCHERICHIA COLI.  
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
 OC ESCHERICHIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K-12;  
 RX MEDLINE: 97426617.  
 RA BLATTNER F.R., PLUNKETT III G., BLOCH C.A., PERNA N.T., BURLAND V.,  
 RA RILEY M., COLLADO-VIDES J., GLASNER J.D., RODE C.K., MAYHEW G.F.,  
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,  
 RA MAU B., SHAO Y.;  
 RT "The complete genome sequence of *Escherichia coli* K-12."  
 RL SCIENCE 277:1453-1474(1997).  
 DR EMBL: AE000137; G1786500; -.  
 SQ SEQUENCE 70 AA; 7854 MW; B754D938 CRC32;

Query Match 37.5%; Score 66; DB 2; Length 70;

Best Local Similarity 53.3%; Pred. No. 7.55e+00; Mismatches 5; Mismatches 2; Indels 0; Gaps 0;

Db 47 QYVQTSYLSQQSSI 61  
 I:: I:: I:: I:: I:: I::  
 QY 10 QYVQTPFLSESNSV 24

RESULT 11  
 ID O87765 PRELIMINARY; PRT: 215 AA.

AC 087765;  
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE PYROLIDONE CARBOXYL PEPTIDASE (EC 3.4.19.1)  
 DE (ACYLAMINOACYL-PEPTIDASE) (ACYLAMINO-ACID RELEASING ENZYME)  
 DE (N-ACYLPEPTIDE HYDROLASE)  
 GN PCP  
 OS LACTOCOCCUS LACTIS SUBSP. CREMORIS.  
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC LACTOCOCCUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 98406035.  
 RA DAVERAN-MINGOT M.L., CAMPO N., RITZENTHALER P., LE BOURGEOIS P.;  
 RA "A natural large chromosomal inversion in *Lactococcus lactis* is  
 RT mediated by homologous recombination between two insertion  
 RT sequences."  
 RL J. BACTERIOL. 180:4834-4842(1998).  
 CC -1- CATALYTIC ACTIVITY: ACYLAMINOACYL-PEPTIDE + H(2)O = ACYLAMINO  
 CC ACID + PEPTIDE.  
 DR EMBL: AJ2233960; E1323722;  
 KW HYDROLASE.  
 SQ SEQUENCE 215 AA; 23519 MW; 004FCDDF CRC32;  
 Query Match 37.5%; Score 66; DB 2; Length 215;  
 Best Local Similarity 31.6%; Pred. No. 7.55e+00;  
 Matches 6; Conservative 8; Mismatches 5; Indels 0; Gaps 0;  
 Db 153 DKMPDPTKAGFMHPIPLPE 171  
 QY 2 EKFMAEFGQGYVOTPFLE 20  
 RESULT 12  
 ID Q47402 PRELIMINARY; PRT; 296 AA.  
 AC Q47402;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE REGION 1 GENE PRODUCT (FRAGMENT).  
 GN KPSS.  
 OS ESCHERICHIA COLI.  
 OC BACTERIA: PROTEOBACTERIA: GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
 OC ESCHERICHIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92138601.  
 RA STEENBERGEN S.M., WRONA T.J., VIMR E.R.;  
 RT "Functional analysis of the sialyltransferase complexes in  
 RT *Escherichia coli* K1 and K92."  
 RL J. BACTERIOL. 174:1099-1108(1992).  
 DR EMBL: M76370; G146950;  
 FT NON-TER  
 SQ SEQUENCE 296 AA; 35733 MW; E10BA47B CRC32;  
 Query Match 37.5%; Score 66; DB 2; Length 296;  
 Best Local Similarity 34.8%; Pred. No. 7.55e+00;  
 Matches 8; Conservative 9; Mismatches 4; Indels 2; Gaps 2;  
 Db 2 RFLA-FREGYLRPOFIVTEEDGV 23  
 QY 3 KFMAEFGQGYVOTPFLE-ESNSV 24  
 RESULT 13  
 ID Q23414 PRELIMINARY; PRT; 334 AA.  
 AC Q23414;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE COSMID ZK1248.  
 GN ZK1248.5.

OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA: METAZOA: NEMATODA: SECERNITEA: RHABDITIA: RHABDITIDA;  
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-BRISTOL N2;  
 RX MEDLINE: 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SVALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
 RA THIERRY-MIES J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*  
 RT *elegans*."  
 RL NATURE 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA LATREILLE P.;  
 RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA WATERSTON R.;  
 RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: U29244; G862494;  
 SQ SEQUENCE 334 AA; 38293 MW; 9ACC97C2 CRC32;  
 Query Match 37.5%; Score 66; DB 5; Length 334;  
 Best Local Similarity 39.1%; Pred. No. 7.55e+00;  
 Matches 9; Conservative 4; Mismatches 9; Indels 1; Gaps 1;  
 Db 55 IKKFKAWGNDYHOTLFLKDS 77  
 QY 1 MEKFMAEFGQGYVOT-PFLSEN 22  
 RESULT 14  
 ID Q58468 PRELIMINARY; PRT; 336 AA.  
 AC Q58468;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)  
 DE 336AA LONG HYPOTHETICAL PROTEIN.  
 GN PH0737.  
 OS PYROCOCCLUS HORIKOSHII.  
 OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCCLUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OT3;  
 RX MEDLINE: 98344137.  
 RA KAWABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,  
 RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOTAMA A., NAGAI Y.,  
 RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,  
 RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,  
 RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,  
 RA KIKUCHI H.;  
 RT "Complete Sequence and Gene Organization of the Genome of a  
 RT Hyper-thermophilic Archaeobacterium, *Pyrococcus horikoshii* OT3."  
 RL DNA RES. 5:55-76(1998).  
 DR EMBL: AF000003; D1030771;  
 SQ SEQUENCE 336 AA; 37579 MW; 62430DEB CRC32;  
 Query Match 37.5%; Score 66; DB 1; Length 336;  
 Best Local Similarity 38.9%; Pred. No. 7.55e+00;  
 Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 Db 166 EYVNGFVKSHFLDDKASV 183





\*\*\*\*\*  
 W P S R L A  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:09:02 1999; MasPar time 3.86 Seconds  
 Tabular output not generated. 93.552 Million cell updates/sec

Title: >PCT-US99-13024-2  
 Description: (1-17) from PCTUS9913024.pcp (3 of 12)  
 Perfect Score: 138  
 Sequence: 1 MEKFWAEFGQGVQPPF 17  
 Scoring table: PAM 150  
 Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-geneseq35  
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21 22:part22 23:part23  
 24:part24 25:part25 26:part26 27:part27 28:part28  
 29:part29 30:part30 31:part31 32:part32 33:part33  
 34:part34 35:part35 36:part36 37:part37 38:part38  
 39:part39

Statistics: Mean 20.181; Variance 71.670; scale 0.282

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	63	45.7	389	2	D-alanine racemase.	4.48e+01
2	60	43.5	130	36	S. pneumoniae protein	8.42e+01
3	59	42.8	38	35	Leader peptide sequen	1.04e+02
4	59	42.8	364	33	Streptococcus pneumon	1.04e+02
5	59	42.8	380	35	Glycerol dehydrogenas	1.04e+02
6	59	42.8	409	9	Sequence of protease	1.04e+02
7	59	42.8	740	33	Streptococcus pneumon	1.04e+02
8	58	42.0	578	23	Human oxalyl-CoA deca	1.28e+02
9	58	42.0	578	14	Oxalyl-CoA decarboxyl	1.28e+02
10	58	42.0	2721	2	Sequence of N-termina	1.28e+02
11	58	42.0	4536	28	Apolipoprotein B-100.	1.28e+02
12	58	42.0	5069	30	A. mediterranei rifam	1.28e+02
13	57	41.3	65	19	Phage T4 ORFX gene pr	1.57e+02
14	57	41.3	321	8	Carnitine dehydrogena	1.57e+02
15	57	41.3	623	36	HSV-2 strain SB5 Cont	1.57e+02
16	57	41.3	1196	36	HSV-2 strain SB5 Cont	1.57e+02

	17	57	41.3	1670 28	W41314	Full length heat-resi	1.57e+02
18	57	41.3	1670 16	R85599	DNA polymerase from s	1.57e+02	
19	57	41.3	1670 18	R94616	DNA polymerase.	1.57e+02	
20	57	41.3	1670 18	R97048	DNA polymerase, contg	1.57e+02	
21	57	41.3	1670 23	W20049	KOD1 thermostable DNA	1.57e+02	
22	56	40.6	43 8	R40095	Hib OMP Pl-P2 hybrid	1.92e+02	
23	56	40.6	461 30	W39928	Human CDP-diacylglyce	1.92e+02	
24	56	40.6	979 39	W83068	TuT protein of tolu	1.92e+02	
25	55	39.9	17 12	R63065	Acetyl-esterase II N-	2.36e+02	
26	55	39.9	218 39	W89435	Streptococcus pneumon	2.36e+02	
27	55	39.9	218 39	W83371	Streptococcus pneumon	2.36e+02	
28	55	39.9	486 22	W20586	H. pylori cytoplasmic	2.36e+02	
29	55	39.9	518 39	W67616	A. nidulans phenylace	2.36e+02	
30	55	39.9	622 8	R38888	Sequence encoded by O	2.89e+02	
31	54	39.1	23 10	R53794	T. litoralis IVPS1 (3	2.89e+02	
32	54	39.1	23 37	W70570	Protein-splice juncti	2.89e+02	
33	54	39.1	344 8	R40918	Bacillus subtilis ino	2.89e+02	
34	54	39.1	458 4	R22279	Human gp.-specific co	2.89e+02	
35	54	39.1	458 4	R22278	Human gp.-specific co	2.89e+02	
36	54	39.1	458 23	W10561	Vitamin D3-binding pr	2.89e+02	
37	54	39.1	458 11	R56975	Variant vitamin D bin	2.89e+02	
38	54	39.1	481 26	W34554	ML17L glycosidase 29G	2.89e+02	
39	54	39.1	621 29	W55645	H. pylori ORF 06epl06	2.89e+02	
40	54	39.1	645 37	W73030	Helicobacter pylori 7	2.89e+02	
41	54	39.1	834 2	R08391	Sequence encoded by v	2.89e+02	
42	54	39.1	911 18	R95634	DNA-ligase-III.	2.89e+02	
43	54	39.1	1312 10	R53787	Deep Vent DNA polymer	2.89e+02	
44	54	39.1	1312 37	W70563	DNA polymerase of a	2.89e+02	
45	54	39.1	1829 27	W29322	DNA polymerase with 3	2.89e+02	

## ALIGNMENTS

RESULT 1  
 ID P70668 standard; protein; 389 AA.  
 AC P70668;  
 DT 11-MAR-1991 (first entry)  
 DE D-alanine racemase.  
 KW L-glutamate racemase; D-alanyl-D-alanine ligase.  
 OS Bacillus subtilis.  
 PN G82177097-A.  
 PD 14-JAN-1987.  
 PF 17-JUN-1986; 014702.  
 PR 18-JUN-1985; US-746437.  
 PA (GENE-) GENENCOR INC.  
 PI Ferrari E;  
 DR WPI; 87-009482/02.  
 DR N-FSD8; N70668.  
 PT Stable maintenance of heterologous DNA in host cell - using host  
 PT cells deficient in ability to synthesise cell wall transformed to  
 PT regain ability  
 PS Disclosure; Fig 2; 13pp; English.  
 CC Protein may be produced in a transformed cell-line deficient in  
 CC D-alanine production. The cell will then express the required sequence  
 CC and a second sequence operably linked to the D-alanine racemase.  
 CC Other sequences which may be used in the same way include those  
 CC encoding L-glutamate racemase or D-alanyl-D-alanine ligase.  
 SQ Sequence 389 AA;  
 Query Match 45.7%; Score 63; DB 2; Length 389;  
 Best Local Similarity 50.0%; Pred. No. 4.48e+01;  
 Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Db 313 mdqfmevdqey 324  
 QY I:|:|:|:|:  
 1 MEKFWAEFGQGV 12  
 RESULT 2  
 ID W80717 standard; Protein; 130 AA.  
 AC W80717;  
 DT 24-DEC-1998 (first entry)  
 DE S. pneumoniae protein of unknown function.

KW Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;  
 KW virulence; antibody; infection; detection; treatment; hypothetical;  
 OS cell wall biosynthetic, external target; minimal gene set protein.  
 PN Streptococcus pneumoniae.  
 PD WO9826072-A1.  
 PF 18-JUN-1998.  
 PR 09-DEC-1997; U22578.  
 PR 13-DEC-1996; US-036281.  
 PA (ELIL) LILLY & CO ELI.  
 PI Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR,  
 PI Mills BJ, Norris FH, Peery RB, Rockey PK, Rostack PR,  
 PI Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ,  
 PI Young Beilido ML;  
 DR WPI; 98-348529/30.  
 DR N-PSDB; V65274.  
 PT Streptococcus pneumoniae nucleic acid sequences - used in DNA chips  
 PT for evaluating gene expression, and identification of virulence  
 PT genes  
 PS Claim 3: Pages 306-307; 333pp; English.  
 CC This sequence represents a Streptococcus pneumoniae protein of unknown  
 CC function. The invention provides DNA sequences (V65201 to V65304) from  
 CC the Streptococcus pneumoniae genome and corresponding protein sequences  
 CC (W60605 to W80728). The protein sequences are classified as hypothetical,  
 CC cell wall biosynthetic, external target, or minimal gene set proteins. A  
 CC recombinant host containing a vector comprising any of the above nucleic  
 CC acids can be used for the recombinant expression of the proteins. The  
 CC invention also provides a DNA chip having arrayed on it at least 15 base  
 CC pair fragment of any one or more of these DNA sequences. The DNA chip can  
 CC be used methods for evaluating gene expression in S. pneumoniae and for  
 CC identifying virulence genes in S. pneumoniae. Antibodies that selectively  
 CC bind to the above proteins or peptide fragments can be used to treat  
 CC S. pneumoniae infection. The antibodies can also be used to detect  
 CC S. pneumoniae cells.  
 SQ Sequence 130 AA;

Query Match 43.5%; Score 60; DB 36; Length 130;  
 Best Local Similarity 53.8%; Pred. No. 8.42e+01;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 69 eqimadlaqvqvq 81  
 QY I : : : : : I I I  
 2 EKFMAEFGQGYVQ 14

RESULT 3  
 ID W78736 standard; peptide; 38 AA.  
 AC W87736;  
 DT 05-NOV-1998 (first entry)  
 DE Leader peptide sequence TA75 50.  
 KW YAP3 signal peptide; synthetic leader peptide; construct; yeast;  
 KW insulin; insulin-like growth factor; glucagon.  
 OS Synthetic.  
 PN WO9832867-A1.  
 PD 30-JUL-1998.  
 PR 22-JAN-1998; DK00026.  
 PR 24-JAN-1997; DK-000097.  
 PA (NOVO) NOVO-NORDISK AS.  
 PI Balschmidt P, Havelund S, Kjellden TB, Pettersson AF;  
 DR WPI; 98-427963/36.  
 PT New DNA construct containing synthetic leader sequence with no  
 PT N-linked glycosylation - for high yield and efficient expression and  
 PT secretion of proteins, e.g. insulin, in yeast, also related vectors  
 PT and transformed yeast  
 PS Claim 14; Page 10; 41pp; English.  
 CC A DNA construct has been developed encoding a polypeptide and has the  
 CC formula given below, components in parentheses may be omitted:  
 CC SP-LP-(PS)-(S)-(PS)-gene (I); where SP = sequence encoding a signal  
 CC peptide; LP = sequence encoding synthetic leader peptide without any  
 CC N-linked glycosylation; PS = sequence encoding a protease processing  
 CC site; S = sequence encoding a spacer peptide; gene = sequence encoding a  
 CC polypeptide. Also described are: (1) an expression cassette (EC)  
 CC comprising (I), fused to a 5'-promoter and 3'-terminator; (2) a yeast  
 CC expression vector containing (I); (3) a yeast cell containing the vector

CC as in (2); (4) DNA encoding a synthetic preproleader sequence that lacks  
 CC the consensus N-linked glycosylation site NXT/S (X = any codable amino  
 CC acid other than P); and (5) the preproleader sequence. Expression vectors  
 CC of (2) are used to produce (both express and secrete) polypeptides in  
 CC yeast, e.g. insulin, insulin-like growth factor, glucagon or its  
 CC fragments. The synthetic preproleader sequences are used to direct  
 CC secretion of proteins from yeast cells. These LP provide high yield and  
 CC more efficient recovery and/or purification of secreted proteins  
 CC expressed in eukaryotic cells. The present sequence represents a  
 CC specifically claimed leader peptide sequence of the present invention.  
 SQ Sequence 38 AA;

Query Match 42.8%; Score 59; DB 35; Length 38;  
 Best Local Similarity 35.3%; Pred. No. 1.04e+02;  
 Matches 6; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

Db 11 mvnlmadddgfmqapl 27  
 QY I : : : : : I : : : : : I  
 1 MEKFMAEFGQGYVQTPF 17

RESULT 4  
 ID W62047 standard; Protein; 364 AA.  
 AC W62047;  
 DT 01-OCT-1998 (first entry)  
 DE Streptococcus pneumoniae spo/rei distal terminal portion.  
 KW Streptococcus pneumoniae; spo/rei; spo/reiA family; screening;  
 KW antibacterial; otitis media; conjunctivitis; pneumonia; bacteraemia;  
 KW meningitis.  
 OS Streptococcus pneumoniae.  
 PN EP-849362-A2.  
 PD 24-JUN-1998.  
 PR 23-OCT-1997; 308473.  
 PR 24-OCT-1996; US-029049.  
 PA (SMIK) SMITHKLINE BEECHAM CORP.  
 PI Gentry DR;  
 PI WPI; 98-324673/29.  
 DR N-PSDB; V37948.  
 PT New Streptococcus spo/reiA polypeptide(s) - useful in treatment of  
 PT otitis media, conjunctivitis, pneumonia, bacteraemia and especially  
 PT meningitis  
 PS Claim 11; Page 31-32; 40pp; English.  
 CC The present sequence is the distal terminal portion of a spo/rei  
 CC protein (spo/reiA family) from Streptococcus pneumoniae. The protein is  
 CC of the spo/reiA family, which, in Escherichia coli, is involved in the  
 CC stringent response to nutrient limitation and regulate the accumulation  
 CC of (p)ppGpp which is involved in the regulation of gene expression and  
 CC other cellular processes. Spo/rei proteins have a homology to e.g.  
 CC S. equismilis rel protein. Spo/rei proteins can be used: (a) in the  
 CC treatment of an individual in need of spo/rei protein, by administering  
 CC to the protein to the patient; (b) in the treatment of an individual  
 CC having need to inhibit spo/rei protein, by administering an antagonist  
 CC which inhibits the activity of the protein; and (c) for inducing an  
 CC immunological response by inoculating the mammal with spo/rei protein,  
 CC or a fragment or variant of it, adequate to produce antibody and/or T  
 CC cell immune response to protect the animal from disease. Conditions  
 CC which may be treated include otitis media, conjunctivitis, pneumonia,  
 CC bacteraemia, meningitis, sinusitis, pleural empyema, endocarditis and  
 CC especially meningitis and compositions may be used as antibacterials.  
 SQ Sequence 364 AA;

Query Match 42.8%; Score 59; DB 33; Length 364;  
 Best Local Similarity 47.1%; Pred. No. 1.04e+02;  
 Matches 8; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

Db 119 emlmaqfgengyankf 135  
 QY I : : : : : I : : : : : I  
 2 EKFMAEFGQ-GYVQTPF 17

RESULT 5  
 ID W69360 standard; Protein; 380 AA.  
 AC W69360;





CC angiogenesis, cellular differentiation or apoptosis. KRAD-14, which is  
CC active as such or as part of a 98-aa peptide, inhibits activation of the  
CC prothrombinase complex; and prevents activation of factor VII on the  
CC surface of thromboplastin and of platelets by thrombin. It binds to the  
CC residues 58-66 of thromboplastin. Since (I) are much smaller than  
CC apoB-100, they act more quickly.  
SQ Sequence 4536 AA;

Query Match 42.0%; Score 58; DB 28; Length 4536;  
Best Local Similarity 23.5%; Pred. No. 1.28e+02;  
Matches 4; Conservative 7; Mismatches 5; Indels 1; Gaps 1;

Db 2147 lqymiqfdq-yikdsy 2162  
::: ||||| ::  
QY 1 MEKFMFEFGQGYVQTPF 17

RESULT 12  
ID W52846 standard; Protein; 5069 AA.  
AC W52846;  
DT 24-JUL-1998 (first entry)  
DE A. mediterranei rifamycin synthesis gene cluster fragment protein B.  
KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;  
OS polyketide synthase; actinomycete; ansamycin.  
PN Amycolatopsis mediterranei.  
PD 26-FEB-1998.  
PF 18-AUG-1997; E04495.  
PR 20-AUG-1996; EP-810551.  
PA (NOVS ) NOVARTIS AG.  
PI Engel N, Schupp T, Toupet C;  
DR WPI; 98-169172/15.  
DR N-PSDB; V21187.

PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used  
PT to produce rifamycin and rifamycin analogues  
PS Claim 6: Page 126-151; 205pp; English.  
CC The present sequence represents a Amycolatopsis mediterranei rifamycin  
CC synthesis gene cluster ORF B protein from the present invention. The  
CC DNA fragment comprises a DNA region involved directly or indirectly  
CC in the gene cluster responsible for rifamycin synthesis, including  
CC the adjacent DNA regions to the right and left which, by reason of  
CC their function in connection with rifamycin biosynthesis, qualify  
CC as constituents of this rifamycin gene cluster, and functional  
CC fragments, derivatives or constituents of these. The Amycolatopsis  
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used  
CC for producing rifamycin, rifamycin analogues or precursors. It can also  
CC be used for inactivating or modifying genes involved in ansamycin or  
CC rifamycin biosynthesis. The DNA can be used for constructing mutant  
CC actinomycetes strains from which the natural rifamycin or ansamycin  
CC biosynthesis gene cluster has been partly or completely deleted. The  
CC DNA fragment can be used for assembling a library of polyketide  
CC synthases, which can be used for assembling a library of polyketide  
CC A hybridisation probe of the invention can be used for identifying DNA  
CC fragments involved in the biosynthesis of ansamycins.  
SQ Sequence 5069 AA;

Query Match 42.0%; Score 58; DB 30; Length 5069;  
Best Local Similarity 70.0%; Pred. No. 1.28e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 3451 ftgmfgggv 3460  
| : |||||  
QY 4 FMAEFGQGYV 13

RESULT 13  
ID R97371 standard; Protein; 65 AA.  
AC R97371;  
DT 07-JAN-1997 (first entry)  
DE Phase T4 ORFX gene product. gp34.  
KW Phase T4; tail fibre protein; nanotechnology; nano-structure;  
KW filter; molecular sieve.  
OS Bacteriophage T4.

PN WO9611947-A1.  
PD 25-APR-1996.  
PF 13-OCT-1995; UI3023.  
PR 13-OCT-1994; US-322760.  
PA (GOLD/) GOLDBERG E B.  
PI Goldberg EB;  
DR WPI; 96-221942/22.  
DR N-PSDB; T29053.

PT New proteins derived from T4 phage tail fibre proteins - that can  
PT self assemble into nano-structure(s), useful as filters etc, also  
PT corresponding DNA

PS Claim 7; Fig 7; 83pp; English.  
CC A protein (R97371) of unspecified function is the product of open  
CC reading frame X of the tail fibre protein gene region (see also  
CC T29053) of phage T4. This gene region also includes open reading  
CC frames for tail fibre proteins (see also R97370 and R97372-74).  
CC Tail fibre proteins (native or modified) can be produced in large  
CC quantities in microbial cells and used as building blocks of strong,  
CC stable nanostructures.  
SQ Sequence 65 AA;

Query Match 41.3%; Score 57; DB 19; Length 65;  
Best Local Similarity 87.5%; Pred. No. 1.57e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 mekfmaei 8  
|||||||  
QY 1 MEKFMFEFG 8

RESULT 14  
ID R36268 standard; Protein; 321 AA.  
AC R36268;  
DT 19-JAN-1994 (first entry)  
DE Carnitine dehydrogenase.  
KW CDHase; clinical measurement.  
OS Alkaligenes sp. NO. 981 FERM BP-2570.  
PN J05161492-A.  
PD 29-JUN-1993.  
PF 08-JUL-1991; 193471.  
PR 08-JUL-1991; JP-193471.  
PA (ASAH ) ASAH CHEM IND CO LTD.  
DR WPI; 93-239257/30.  
DR N-PSDB; Q46585.

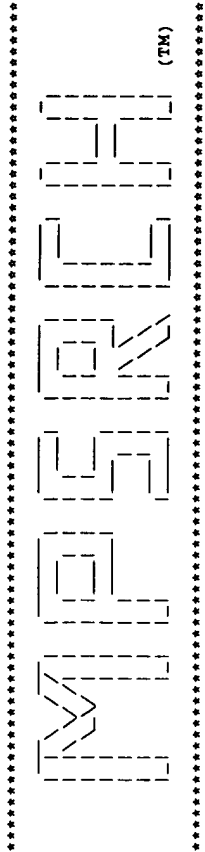
PT New purified Escherichia coli producing carnitine dehydrogenase -  
PT useful for clinical measurement of carnitine  
PS Disclosure; Pages 11-13; 21pp; Japanese.  
CC The sequence is that of carnitine dehydrogenase (CDHase) which  
CC may be produced in E. coli transformed with an expression plasmid  
CC containing the CDHase gene without the addition of carnitine. CDHase  
CC is useful for clinical measurement of carnitine.  
SQ Sequence 321 AA;

Query Match 41.3%; Score 57; DB 8; Length 321;  
Best Local Similarity 50.0%; Pred. No. 1.57e+02;  
Matches 8; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

Db 246 mrhfmaeqfpa-lqlp 260  
| ||||| : ||  
QY 1 MEKFMFEFGQGYVQTP 16

RESULT 15  
ID W72197 standard; Protein; 623 AA.  
AC W72197;  
DT 13-JAN-1999 (first entry)  
DE HSV-2 strain SB5 Contig ID 15 ORF#32 protein.  
KW HSV-2 strain SB5; immunological response induction; therapy;  
KW antiviral identification; viral protein inhibitor.  
OS Herpes simplex virus type 2.  
PN WO9820016-A1.  
PD 14-MAY-1998.  
PF 31-OCT-1997; U20016.





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:08:32 1999; MasPar time 3.94 Seconds  
Tabular output not generated.  
172.832 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (1-17) from PCTUS9913024.pep (3 of 12)  
Perfect Score: 138  
Sequence: 1 MEKFAEFGQGVQTFP 17

Scoring table: PAM 150  
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 28.908; Variance 46.477; scale 0.622

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description	Pred. No.
1	77	55.8	860	2	F71000	hypothetical protein
2	66	47.8	56	2	S35168	cytochrome P450 (clon
3	66	47.8	508	2	I64032	hypothetical protein
4	65	47.1	2301	2	T02323	hypothetical protein
5	64	46.4	273	2	D70428	hypothetical protein
6	64	46.4	420	1	TVUTG4	phosphoglycerate kin
7	64	46.4	440	1	KIUTGC	phosphoglycerate kin
8	64	46.4	440	1	TVUTGB	phosphoglycerate kin
9	64	46.4	505	1	TVUT2B	phosphoglycerate kin
10	64	46.4	508	1	TVUT4B	phosphoglycerate kin
11	64	46.4	509	1	AA5593	phosphoglycerate kin
12	63	45.7	389	1	JS0443	alanine racemase (EC
13	63	45.7	417	1	TVCRGC	phosphoglycerate kin
14	63	45.7	455	1	TVCRGG	phosphoglycerate kin
15	63	45.7	455	2	E71569	probable acyltransfer
16	63	45.7	603	1	W1WL8	El protein - human pa
17	63	45.7	645	2	F70825	probable PPE protein
18	63	45.7	1215	2	S69004	hypothetical protein
19	62	44.9	191	2	C69054	fuculose-1-phosphate
20	62	44.9	421	1	KIUTGC	phosphoglycerate kin
21	62	44.9	858	2	S68227	ubiquitin thiolester
22	61	44.2	605	2	S36592	El protein - human pa
23	60	43.5	205	2	E69327	Ribonuclease H1 (rnh

24	60	43.5	558	2	F71175	hypothetical protein
25	60	43.5	754	2	S52816	probable membrane pro
26	60	43.5	758	2	I64084	hemoglobin receptor h
27	60	43.5	1394	2	S66876	ATP-dependent transpo
28	59	42.8	163	2	S49633	hypothetical protein
29	59	42.8	262	2	S64050	hypothetical protein
30	59	42.8	380	2	D65201	glycerol dehydrogenas
31	59	42.8	484	1	A24994	cellulose 1,4-beta-ce
32	59	42.8	609	2	S36481	El protein - human pa
33	59	42.8	610	2	JH0573	chitinase (EC 3.2.1.1
34	59	42.8	681	2	S36534	El protein - human pa
35	59	42.8	762	2	A54411	amine oxidase (copper
36	59	42.8	809	1	Q0BE34	BBLF4 protein - human
37	59	42.8	884	2	A31928	glucose transport pro
38	59	42.8	936	2	H71862	probable cytochrome C
39	59	42.8	936	2	B64567	cytochrome c biogenes
40	59	42.8	1007	2	G70374	hypothetical protein
41	59	42.8	1465	2	A70199	hypothetical protein
42	58	42.0	365	2	D69682	gamma-glutamyl kinase
43	58	42.0	604	2	S36493	El protein - human pa
44	58	42.0	605	2	S36469	El protein - human pa
45	58	42.0	4563	1	LPHUB	apolipoprotein B-100

ALIGNMENTS

RESULT 1

ENTRY #type complete  
TITLE hypothetical protein PH1302 - Pyrococcus horikoshii  
ORGANISM #formal\_name Pyrococcus horikoshii  
DATE 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 14-Aug-1998

ACCESSIONS F71000  
REFERENCE A71000

#authors Kwarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.

#journal DNA Res. (1998) 5:55-76  
#title Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.

#cross-references MUID:98344137  
#accession F71000  
#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule\_type DNA  
#residues 1-860 #label KAW  
#cross-references GB:AP000006; NID:g3236133; PID:d1031349; PID:g3257723  
#experimental\_source strain OT3  
#note this accession replaces an interim accession for a this sequence replaced by GenBank

GENETICS PH1302  
SUMMARY #length 860 #molecular-weight 95454 #checksum 3711

Query Match 55.8%; Score 77; DB 2; Length 860;  
Best Local Similarity 29.4%; Pred. No. 3.00e-02;  
Matches 5; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

Db 1 MSKFLTDFSNATRVAF 17  
QY 1 MEKFAEFGQGVQTFP 17  
| | | | | : : : : : |

RESULT 2  
ENTRY S35168 #type fragment  
TITLE cytochrome P450 (Clone 3) - Madagascar periwinkle (fragment)  
CONTAINS oxidoreductase (EC 1.-.-.-)  
ORGANISM #formal\_name Catharanthus roseus #common\_name Madagascar









```
CLASSIFICATION #superfamily alanine racemase
KEYWORDS cell wall synthesis; isomerase; phosphoprotein; pyridoxal
          phosphate
FEATURE
41 #binding_site pyridoxal phosphate (Lys) (covalent)
SUMMARY #length 389 #molecular-weight 43327 #checksum 9391

Query Match 45.7%; Score 63; DB 1; Length 389;
Best Local Similarity 50.0%; Pred. No. 6.76e+00;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 313 MDQWVLDQY 324
| : ||| :| :| :|
Oy 1 MEKFMAEFGQGYVQTPF 12

RESULT 13
ENTRY TVCRGG #type complete
TITLE phosphoglycerate kinase (EC 2.7.2.3), cytosolic - Crithidia
ALTERNATE_NAMES fasciculata
ORGANISM phosphoglycerate kinase B
#formal_name Crithidia fasciculata
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
05-Sep-1997
ACCESSIONS S00486
REFERENCE Swinkels, B.W.; Evers, R.; Borst, P.
EMBO J. (1988) 7:1159-1165
#journal The topogenic signal of the glycosomal (microbody)
#title phosphoglycerate kinase of Crithidia fasciculata resides in
a carboxy-terminal extension.
#cross-references MUID:88296420
#accession S00486
#molecule_type DNA
#residues 1-417 #label SWI
#cross-references EMBL:X07458; NID:g6982; PID:g6983
#note the authors translated the codon GAC for residue 173 as
293 as Ala

CLASSIFICATION #superfamily phosphoglycerate kinase
KEYWORDS ATP; gluconeogenesis; glycolysis; phosphotransferase
FEATURE
220,342 #binding_site ATP (Lys, Glu) #status predicted
SUMMARY #length 417 #molecular-weight 44602 #checksum 3677

Query Match 45.7%; Score 63; DB 1; Length 417;
Best Local Similarity 41.2%; Pred. No. 6.76e+00;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Db 148 MAKVLAAYGDVYVSDAF 164
| :| :| :| :| :|
Oy 1 MEKFMAEFGQGYVQTPF 17

RESULT 14
ENTRY TVCRGG #type complete
TITLE phosphoglycerate kinase (EC 2.7.2.3), glycosomal - Crithidia
ALTERNATE_NAMES fasciculata
ORGANISM phosphoglycerate kinase C
#formal_name Crithidia fasciculata
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
05-Sep-1997
ACCESSIONS S00487
REFERENCE Swinkels, B.W.; Evers, R.; Borst, P.
EMBO J. (1988) 7:1159-1165
#journal The topogenic signal of the glycosomal (microbody)
#title phosphoglycerate kinase of Crithidia fasciculata resides in
a carboxy-terminal extension.
#cross-references MUID:88296420
#accession S00487
#molecule_type DNA
```

```
#residues 1-455 #label SWI
#cross-references EMBL:X07459; NID:g6984; PID:g6985
#note the authors translated the codon GAC for residue 173 as
Thr, CAA for residue 292 as Lys, and GGC for residue
293 as Ala
CLASSIFICATION #superfamily phosphoglycerate kinase
KEYWORDS ATP; glycolysis; glycosome; phosphotransferase
FEATURE
220,342 #binding_site ATP (Lys, Glu) #status predicted
SUMMARY #length 455 #molecular-weight 47843 #checksum 8443

Query Match 45.7%; Score 63; DB 1; Length 455;
Best Local Similarity 41.2%; Pred. No. 6.76e+00;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Db 148 MAKVLAAYGDVYVSDAF 164
| :| :| :| :| :|
Oy 1 MEKFMAEFGQGYVQTPF 17

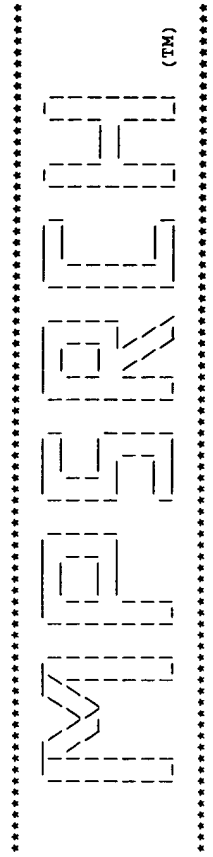
RESULT 15
ENTRY E71569 #type complete
TITLE Probable acyltransferase - Chlamydia trachomatis (serotype D,
strain UM3/Cx)
ORGANISM #formal_name Chlamydia trachomatis
DATE 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change
21-Nov-1998
ACCESSIONS E71569
REFERENCE A71570
#authors Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe,
R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov,
R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
#journal Science (1998) 282:754-759
#title Genome sequence of an obligate intracellular pathogen of
humans: Chlamydia trachomatis.
#cross-references MUID:99000809
#accession E71569
#status preliminary
#molecule_type DNA
#residues 1-455 #label ARN
#cross-references GB:AE001275; GB:AE001273; NID:g3328388; PID:g3328398
#experimental_source serotype D, strain UM-3/Cx

GENETICS
#gene htrB
SUMMARY #length 455 #molecular-weight 52058 #checksum 3772

Query Match 45.7%; Score 63; DB 2; Length 455;
Best Local Similarity 29.4%; Pred. No. 6.76e+00;
Matches 5; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

Db 343 LQRFLIEFGFYADASL 359
| :| :| :| :| :| :|
Oy 1 MEKFMAEFGQGYVQTPF 17

Search completed: Wed Sep 1 16:08:44 1999
Job time : 12 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:07:26 1999; Maspar time 2.71 Seconds  
Tabular output not generated. 177.194 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (1-17) from PCTUS9913024.pap (3 of 12)  
Perfect Score: 138  
Sequence: 1 MEKFAEFGQGYVTPF 17

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 29.668; Variance 42.415; scale 0.699

\* Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result	Query	Match	Length	ID	Description	Pred.	No.
1	66	47.8	508	1	YF00_HAEIN HYPOTHETICAL PROTEIN H	7.70e-01	
2	66	47.8	524	1	CP72_CATRO CYTOCHROME P450 72A1 (	7.70e-01	
3	64	46.4	420	1	PGKE_TRYBB PHOSPHOGLYCERATE KINAS	1.73e+00	
4	64	46.4	440	1	PGKC_TRYBB PHOSPHOGLYCERATE KINAS	1.73e+00	
5	64	46.4	505	1	PGKA_TRYBB PHOSPHOGLYCERATE KINAS	1.73e+00	
6	64	46.4	508	1	PGKD_TRYBB PHOSPHOGLYCERATE KINAS	1.73e+00	
7	63	45.7	389	1	ALR_BACSU ALANINE RACEMASE (EC 5	2.59e+00	
8	63	45.7	417	1	PGKB_CRIFA PHOSPHOGLYCERATE KINAS	2.59e+00	
9	63	45.7	455	1	PGKC_CRIFA PHOSPHOGLYCERATE KINAS	2.59e+00	
10	63	45.7	603	1	VEL_HPV08 REPLICATION PROTEIN E1	2.59e+00	
11	62	44.9	417	1	PGKB_LEIMA PHOSPHOGLYCERATE KINAS	3.84e+00	
12	62	44.9	421	1	PGKB_TRYBB PHOSPHOGLYCERATE KINAS	3.84e+00	
13	62	44.9	479	1	PGKC_LEIMA PHOSPHOGLYCERATE KINAS	3.84e+00	
14	62	44.9	604	1	VEL_HPV36 REPLICATION PROTEIN E1	3.84e+00	
15	62	44.9	858	1	UBP5_HUMAN UBIQUITIN CARBOXYL-TER	3.84e+00	
16	62	44.9	858	1	UBP5_MOUSE UBIQUITIN CARBOXYL-TER	3.84e+00	
17	61	44.2	476	1	VTDB_RABIT VITAMIN D-BINDING PROT	5.67e+00	
18	61	44.2	605	1	VEL_HPV09 REPLICATION PROTEIN E1	5.67e+00	
19	60	43.5	734	1	SULX_YEAST PUTATIVE SULFATE TRANS	8.34e+00	
20	60	43.5	1010	1	Y661_HAEIN PROBABLE TONB-DEPENDEN	8.34e+00	
21	59	42.8	163	1	YMJ4_YEAST HYPOTHETICAL 18.4 KD P	1.22e+01	
22	59	42.8	262	1	YGE6_YEAST HYPOTHETICAL 29.4 KD P	1.22e+01	
23	59	42.8	367	1	GLDA_ECOLI GLYCEROL DEHYDROGENASE	1.22e+01	

RESULT	1	484	42.8	59	42.8	484	1	GUX_CELFI	EXOGLUCANASE PRECURSOR	1.22e-01
AC	59	42.8	609	1	VEL_HPV17	CHITINASE 63 PRECURSOR	1.22e-01			
DT	59	42.8	610	1	CHIT_STRPL	CHITINASE C PRECURSOR	1.22e-01			
DT	59	42.8	619	1	CHIT_STRPL	CHITINASE C PRECURSOR	1.22e-01			
27	59	42.8	619	1	CHIT_STRPL	CHITINASE C PRECURSOR	1.22e-01			
28	59	42.8	660	1	VEL_HPV29	REPLICATION PROTEIN E1	1.22e-01			
29	59	42.8	681	1	VEL_HPV10	REPLICATION PROTEIN E1	1.22e-01			
30	59	42.8	762	1	AM02_BOVIN	COPPER AMINE OXIDASE P	1.22e-01			
31	59	42.8	809	1	HELI_EBV	PROBABLE HELICASE	1.22e-01			
32	59	42.8	818	1	SNF3_YEAST	HIGH-AFFINITY GLUCOSE	1.22e-01			
33	59	42.8	809	1	HELI_EBV	PROBABLE HELICASE	1.22e-01			
34	58	42.0	408	1	ALR_MYCTU	GLUTAMATE 5-KINASE (EC 5	1.78e-01			
35	58	42.0	497	1	DYJ2_RAT	ALANINE RACEMASE (EC 5	1.78e-01			
36	58	42.0	604	1	VEL_HPV25	REPLICATION PROTEIN E1	1.78e-01			
37	58	42.0	604	1	VEL_HPV19	REPLICATION PROTEIN E1	1.78e-01			
38	58	42.0	605	1	VEL_HPV14	REPLICATION PROTEIN E1	1.78e-01			
39	58	42.0	606	1	VEL_HPV7	REPLICATION PROTEIN E1	1.78e-01			
40	58	42.0	632	1	ETFD_SCHPO	PROBABLE ELECTRON TRAN	1.78e-01			
41	58	42.0	4563	1	APB_HUMAN	APOLIPROTEIN B-100 P	1.78e-01			
42	57	41.3	314	1	PYRD_YEAST	DIHYDROOCTATE DEHYDRO	2.57e-01			
43	57	41.3	421	1	Y236_METJA	HYPOTHETICAL PROTEIN M	2.57e-01			
44	57	41.3	837	1	XYNZ_CLOTM	ENDO-1,4-BETA-XYLANASE	2.57e-01			
45	57	41.3	2009	1	SEC7_YEAST	PROTEIN TRANSPORT PROT	2.57e-01			

ALIGNMENTS

ID	YF00_HAEIN	STANDARD;	PRT;	508 AA.
AC	P44224;			
DT	01-NOV-1995 (REL. 32, CREATED)			
DT	01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)			
DT	01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)			
DE	HYPOTHETICAL PROTEIN HI1500.			
GN	HI1500.			
OS	HAEMOPHILUS INFLUENZAE.			
OC	BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;			
OC	HAEMOPHILUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-RD / KW20;			
RX	MEDLINE: 95350630.			
RA	FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,			
RA	KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,			
RA	MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,			
RA	SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,			
RA	WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,			
RA	UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,			
RA	FINE L.D., FRITZMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,			
RA	GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,			
RA	VENTER J.C.;			
RT	"Whole-genome random sequencing and assembly of Haemophilus			
RT	Influenzae Rd.;"			
RL	SCIENCE 269:496-512(1995).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL; U32826; G1574341; "			
DR	TIGR; HI1500; "			
KW	HYPOTHETICAL PROTEIN.			
SQ	SEQUENCE 508 AA; 57190 MW; 98DDC6AA CRC32;			
Query Match 47.8%; Score 66; DB 1; Length 508;				
Best Local Similarity 43.8%; Pred. No. 7.70e-01;				
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;				
Db	411 MPKYLKALYESGYIQIP 426			
	:: : :: :			

Qy 1 MEKFAEFGGQYVOTP 16

```
RESULT 2
ID CP72_CATRO STANDARD; PRT; 524 AA.
AC Q05047;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 72A1 (EC 1.14.14.1) (CYPLXXII) (PROBABLE GERANIOL-10-
DE HYDROXYLASE) (GEL0H).
GN CP72A1 OR CYP72 OR P450C83.
OS CATHARANTHUS ROSEUS (ROSY PERIWINKLE) (MADAGASCAR PERIWINKLE).
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTES; TRACHEOPHYTES;
OC EUPHYLOPHYTES; SPERMATOPHYTES; MAGNOLIOPHYTES; EUDICOTYLEDONS;
OC ASTERIDAE; GENTIANACEAE; GENTIANACEAE; APOCYNACEAE; CATHARANTHUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LINE CP3A;
RA VETTER H.-P., MANGOLD U., SCHROEDER G., MARNER F.-J.,
RA WERCK-REICHART D., SCHROEDER J.;
RT "Molecular analysis and heterologous expression of an inducible
RT cytochrome P-450 protein from periwinkle (Catharanthus roseus L.).";
RL PLANT PHYSIOL. 100:998-1007(1992).
RN [2]
RP SEQUENCE OF 469-524 FROM N.A.
RX MEDLINE; 93283641.
RA MEIJER A.H., SOUVER E., VERPOORTE R., HOGE J.H.C.;
RT "Isolation of cytochrome P-450 cDNA clones from the higher plant
RT Catharanthus roseus by a PCR strategy.";
RL PLANT MOL. BIOL. 22:379-383(1993).
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. MAY BE A GERANIOL-10-HYDROXYLASE.
CC -!- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O.
CC -!- PATHWAY: INDOLE ALKALOID SYNTHESIS.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L10081; G167484; -
CC EMBL; X69775; G395302; -
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC PFAM; PF00067; P450; 1.
CC KX OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; TRANSMEMBRANE;
CC HEME.
CC TRANSMEM 12 32 POTENTIAL.
CC BINDING 470 470 HEME (BY SIMILARITY).
CC VARIANT 190 190 I -> L.
CC VARIANT 194 194 Q -> E.
CC VARIANT 223 223 E -> D.
CC VARIANT 312 312 K -> R.
CC VARIANT 318 318 S -> T.
CC VARIANT 403 403 V -> I.
CC VARIANT 405 405 K -> E.
CC VARIANT 411 411 S -> P.
CC SEQUENCE 524 AA; 60557 MW; FC0286BC CRC32;
Query Match 47.8%; Score 66; DB 1; Length 524;
Best Local Similarity 29.4%; Pred. No. 7,70e-01;
Matches 5; Conservative 9; Mismatches 3; Indels 0; Gaps 0;
```

Db 488 LQRFKFDVAPSYVHAPF 504

Qy 1 MEKFAEFGGQYVOTPF 17

```
RESULT 3
ID PGKC_TRYBB STANDARD; PRT; 420 AA.
AC P08893;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE PHOSPHOGLYCERATE KINASE, CYTOSOLIC (B) (EC 2.7.2.3) (ALLELE 4).
OS TRYPANOSOMA BRUCEI BRUCEI.
OC EUKARYOTA; EULENZOEA; KINETOPLASTIDA; TRYPANOSOMATIDAE; TRYPANOSOMA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88286728.
RA LE BLANCQ S.M., SWINKELS B.W., GIBSON W.C., BORST P.;
RT "Evidence for gene conversion between the phosphoglycerate kinase
RT genes of Trypanosoma brucei.";
RL J. MOL. BIOL. 200:439-447(1988).
CC -!- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE -> ADP +
CC 3-PHOSPHO-D-GLYCEROYL PHOSPHATE.
CC -!- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- IN T. BRUCEI, THREE GENES CODE FOR PHOSPHOGLYCERATE KINASE
CC ISOZYMES, WHICH ARE TRANSPORTED TO DIFFERENT CELL COMPARTMENTS.
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X05890; G10494; -
CC PIR; S02334; TVUTG4.
CC PROSITE; PS00111; PGLYCERATE_KINASE; 1.
CC PFAM; PF00162; PGK; 1.
CC DR HSP; P07378; 13PK.
CC KW TRANSPERASE; KINASE; GLYCOLYSIS; MULTIGENE FAMILY.
CC SQ SEQUENCE 420 AA; 45087 MW; DDF3B80 CRC32;
Query Match 46.4%; Score 64; DB 1; Length 420;
Best Local Similarity 35.3%; Pred. No. 1.73e+00;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
Db 151 MAKILASGVGVYISDAF 167
Qy 1 MEKFAEFGGQYVOTPF 17
RESULT 4
ID PGKC_TRYBB STANDARD; PRT; 440 AA.
AC P07378;
DT 01-APR-1988 (REL. 07, CREATED)
DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PHOSPHOGLYCERATE KINASE, GLYCOSOMAL (C) (EC 2.7.2.3).
OS TRYPANOSOMA BRUCEI BRUCEI.
OC EUKARYOTA; EULENZOEA; KINETOPLASTIDA; TRYPANOSOMATIDAE; TRYPANOSOMA.
RN [1]
RP SEQUENCE FROM N.A. (ALLELES 2 AND 4).
RX MEDLINE; 88286728.
RA LE BLANCQ S.M., SWINKELS B.W., GIBSON W.C., BORST P.;
RT "Evidence for gene conversion between the phosphoglycerate kinase
RT genes of Trypanosoma brucei.";
RL J. MOL. BIOL. 200:439-447(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86136022.
RA OSINGA K.A., SWINKELS B.W., GIBSON W.C., BORST P., VEENEMAN G.H.,
RA VAN BOOM J.H., MICHELIS P.A.M., OPPERDOES F.R.;
RT "Topogenesis of microbody enzymes: a sequence comparison of the genes
```

```

J. MOL. BIOL. 200:439-447(1988)).
CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE = ADP +
CC 3-PHOSPHO-D-GLYCEROYL PHOSPHATE.
CC -1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: NOT KNOWN.
CC -1- IN T. BRUCEI, THREE GENES CODE FOR PHOSPHOGLYCERATE KINASE
CC ISOZYMES, WHICH ARE TRANSPORTED TO DIFFERENT CELL COMPARTMENTS.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X05889; G10489; -.
CC PIR; S00748; TVUT28.
CC DR PROSITE; PS00111; GLYCERATE_KINASE; 1.
CC DR PFAM; PF00162; PGK; 1.
CC DR HSP; P07378; 13PK.
CC DR TRANSFERASE; KINASE; GLYCOLYSIS; MULTIGENE FAMILY.
CC SQ SEQUENCE 505 AA; 55376 MW; 92ACCO9 CRC32;
-----
Query Match 46 48; Score 64; DB 1: Length 505;
Best Local Similarity 35.38; Pred. No. 1.73e+00;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 240 MAKILASYGDVYISDAF 256
QY 1 MEKFMFEGGYVQTF 17
| | : | : | : | : |
| | : | : | : | : |

RESULT 6
ID PKGD_TRYBB STANDARD; PRT: 508 AA.
CC P08892;
CC 01-NOV-1988 (REL. 09, CREATED)
CC 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
CC 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
CC DE PHOSPHOGLYCERATE KINASE, A (EC 2.7.2.3) (ALLELE 4).
CC OS TRYPANOSOMA BRUCEI BRUCEI.
CC EUKARYOTA; EUGLENZOZA; KINETOPLASTIDA; TRYPANOSOMATIDAE; TRYPANOSOMA.
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE; 88286728.
CC RX.
CC LE BLANCO S.M., SWINKELS B.W., GIBSON W.C., BORST P.;
CC "Evidence for gene conversion between the phosphoglycerate kinase
CC genes of Trypanosoma brucei.";
CC J. MOL. BIOL. 200:439-447(1988).
CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE = ADP +
CC 3-PHOSPHO-D-GLYCEROYL PHOSPHATE.
CC -1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: NOT KNOWN.
CC -1- IN T. BRUCEI, THREE GENES CODE FOR PHOSPHOGLYCERATE KINASE
CC ISOZYMES, WHICH ARE TRANSPORTED TO DIFFERENT CELL COMPARTMENTS.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X05890; G10493; -.
CC PIR; S02233; TVUT48.
CC DR PROSITE; PS00111; GLYCERATE_KINASE; 1.
CC DR PFAM; PF00162; PGK; 1.
CC DR HSP; P07378; 13PK.

```

KW TRANSFERASE; KINASE; GLYCOLYSIS; MULTIGENE FAMILY.  
SQ SEQUENCE 508 AA; 55668 MW; 7EB9B5FD CRC32;

Query Match 46.4%; Score 64; DB 1; Length 508;  
Best Local Similarity 35.3%; Pred. No. 1.73e+00;  
Matches 6; Conservative 6; Mismatches 5; Indels

```

Db      240 MAKILASYGDVYISDAF 256
      | | : | : | : |
QY      1 MEKFMAEFGQYVQTPE 17

```

RESULT	7	STANDARD:	PRT:	389 AA.
ID	ALR_BACSU			
AC	P10725; P96620;			
DT	01-JUL-1989 (REL. 11, CREATED)			
DT	13-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)			
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)			
DE	ALANINE RACEMASE (EC 5.1.1.1).			
GN	ALR OR DAL.			
OS	BACILLUS SUBTILIS.			
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;			
OC	BACILLUS.			
OC	BACILLUS.			

RT "Isolation of an alanine racemase gene from *Bacillus subtilis* and its  
 RL use for plasmid maintenance in *B. subtilis*."; BIOTECHNOLOGY 3:1003-1007(1985).

```

[2]
RN
RP
RC SEQUENCE FROM N.A.
RC STRAIN=168;
RC
RL KASAHARA Y., NAKAI S., LEE S., SADATE Y., OGASAWARA N.;
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RL
CC -! CATALYTIC ACTIVITY: L-ALANINE - D-ALANINE.
CC -! COFACTOR: PYRIDOXAL PHOSPHATE.
CC -! SUBUNIT: HOMODIMER.
CC -! SIMILARITY: BELONGS TO THE ALANINE RACEMASE FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC

```

```
Query Match      45.7%; Score 63; DB 1; Length 389;
Best Local Similarity 50.0%; Pred. No. 2.59e+00;
Matches      6; Conservative      2; Mismatches 4; Indels 0; Gaps 0;
```

```

Db      313 MDQFMVELDQY 324
      1: 111: 11
OY      1 MEKFMAEFGGY 12

```

RESULT 8  
ID PGKB\_CRIFA STANDARD; PRT; 417 AA.  
AC P08966;

01-NOV-1988 (REL. 09, CREATED)  
01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)  
15-DEC-1988 (REL. 37, LAST ANNOTATION UPDATE)  
PHOSPHOGLYCERATE KINASE, CYTOSOLIC (B) (EC 2.7.2.3).  
PGKB OR PGK-B.  
CRITHIDIA FASCICULATA.  
OS  
EUKARYOTA; EULENZOEA; KINETOPLASTIDA; TRYPANOSOMATIDAE; CRITHIDIA.  
[ ]  
SEQUENCE FROM N.A.  
RP  
MEDLINE; 88296420.  
RX  
SWINKELS B.W., EVERS R., BORST P.;  
RT "The topogenic signal of the glycosomal (microbody) phosphoglycerate  
RT kinase of Crithidia fasciculata resides in a carboxy-terminal  
RT extension".;  
EMBO J. 7:1159-1165(1988).  
RL  
-1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE = ADP +  
CC 3-PHOSPHO-D-GLYCEROYL PHOSPHATE.  
CC  
-1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.  
CC  
-1- SUBUNIT: MONOMER.  
CC  
-1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC  
-1- DOMAIN: THIS CYTOSOLIC PKK LACKS A C-TERMINAL EXTENSION OF  
CC 38 AA WHICH IS PRESENT IN THE GLYCOSOMAL ISOENZYME.  
CC  
-1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.  
CC  
-----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announcement/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

CC -----  
ENBL; X07458: G6983; -  
DR  
DR PIR; S00486; TVCRGC.  
DR PROSITE; PS0011; PGLYCERATE\_KINASE; 1.  
DR PFAM; PF00162; PGK; 1.  
DR HSP; P07378; 13PK.  
DR KW TRANSFERASE; KINASE; GLYCOLYSIS; MULTIGENE FAMILY.  
SQ SEQUENCE 417 AA; 46402 MW; 89E19A31 CRC32;  
CC -----

Query Match 45.7%; Score 63; DB 1; Length 417;  
Best Local Similarity 41.2%; Pred. No. 2.59e+00;  
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Db 148 MAKVLAA YGDVYVSDAF 164  
| | : | : | : | : |  
QY 1 MEKFMAEFGGQYVOTPF 17

RESULT	9	STANDARD:	PRT:	455 AA.
ID	PGKC-CRIFA			
AC	P08967;			
DT	01-NOV-1988 (REL. 09, CREATED)			
DT	01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)			
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)			
DE	DE PHOSPHOGLYCERATE KINASE, GLYCOSOMAL (C) (EC 2.7.2.3.)			
GN	PGKC OR PGR-C.			
OS	CRITHIDIA FASCICULATA.			
OC	EUKARYOTA; EULENZOEA; KINETOPLASTIDA; TRYPANOSOMATIDA			
RN	[1]			

CC - ! SUBCELLULAR LOCATION: GLYCOSOMAL.  
CC - !- DOMAIN: THIS GLYCOSOMAL PGK HAS A C-TERMINAL EXTENSION OF 38 AA

```

Best Local Similarity 50.0%; Pred. No. 2.59e+00;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 166 MAXFKEAFGDF 177
   | | | | | | |
QY 1 MEXFMAEFGQY 12

RESULT 11
ID PKGB_LEIMA STANDARD; PRT: 417 AA.
AC Q27683;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PHOSPHOGLYCERATE KINASE, CYTOSOLIC (B) (EC 2.7.2.3).
GN PKGB OR CPCK.
OS LEISHMANIA MAJOR.
OC EUKARYOTA; EUKLENOZOA; KINETOPLASTIDA; TRYPANOSOMATIDAE; LEISHMANIA.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-ISOLATE FRIEDLIN, LONDON;
RA MCKOY G.E.M., HART D.T.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE = ADP +
CC 3-PHOSPHO-D-GLYCEROYL PHOSPHATE.
CC -1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- DOMAIN: THIS CYTOSOLIC PKG LACKS A C-TERMINAL EXTENSION OF 38 AA
CC WHICH IS PRESENT IN THE GLYCOSOMAL ISOENZYME.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L25120; G499641; -.
DR PROSITE; PS00111; PGLYCERATE_KINASE; 1.
DR PFAM; PF00162; PKG; 1.
DR HSP; P07378; 13PK.
DR KW TRANSFERASE; KINASE; GLYCOLYSIS; MULTIGENE FAMILY.
DR SQ SEQUENCE 417 AA; 44963 MW; B1745DA2 CRC32;

Query Match 44.9%; Score 62; DB 1; Length 417;
Best Local Similarity 41.2%; Pred. No. 3.84e+00;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Db 148 MAXVLASYGDVYVSDF 164
   | | | | | | |
QY 1 MEXFMAEFGQYVQTF 17

RESULT 12
ID PKGB_TRYBB STANDARD; PRT: 421 AA.
AC P07377;
DT 01-APR-1988 (REL. 07, CREATED)
DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
DE 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
DE PHOSPHOGLYCERATE KINASE, CYTOSOLIC (B) (EC 2.7.2.3) (ALLELE 2).
OS TRYPANOSOMA BRUCEI BRUCEI.
OC EUKARYOTA; EUKLENOZOA; KINETOPLASTIDA; TRYPANOSOMATIDAE; TRYPANOSOMA.
RN [1]
RN SEQUENCE FROM N.A.
RC MEDLINE; 88286728.
RA LE BLANCQ S.M., SWINKELS B.W., GIBSON W.C., BORST P.;
RT "Evidence for gene conversion between the phosphoglycerate kinase
RT genes of Trypanosoma brucei.";
RL J. MOL. BIOL. 200:439-447(1988).
RN [2]

```





OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE; 96105388.  
RA FALQUET L., PAQUET N., FRUTIGER S., HUGHES G.J., HOANG-VAN K.,  
RA JATON J.-C.;  
RT "cDNA cloning of a human 100 kDa de-ubiquitinating enzyme: the 100  
RT kDa human de-ubiquitinase belongs to the ubiquitin C-terminal  
RT hydrolase family 2 (UCH2).";  
RL FEBS LETT. 376:233-237(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 96303695.  
RA ANSARI-LARI M.A., MUZYNY D.M., LU J., LU F., LILLEY C.E., SPANOS S.,  
RA MALLEY T., GIBBS R.A.;  
RT "A gene-rich cluster between the CD4 and triosephosphate isomerase  
RT genes at human chromosome 12p13.";  
RL GENOME RES. 6:314-326(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA TASHAYEV V.L., O'CONNOR L.B., LARSEN C.N., KASPEREK E., PICKART C.M.;  
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE; 95154450.  
RA FALQUET L., PAQUET N., FRUTIGER S., HUGHES G.J., HOANG-VAN K.,  
RA JATON J.-C.;  
RT "A human de-ubiquitinating enzyme with both isopeptidase and  
RT peptidase activities in vitro.";  
RL FEBS LETT. 359:73-77(1995).  
CC -!- FUNCTION: CLEAVES LINEAR AND BRANCHED MULTIUBIQUITIN POLYMERS  
CC -!- WITH A MARKED PREFERENCE FOR BRANCHED POLYMERS.  
CC -!- CATALYTIC ACTIVITY: UBIQUITIN C-TERMINAL THIOLESTER + H(2)O -  
CC UBIQUITIN + A THIOL.  
CC -!- COFACTOR: ZINC; REQUIRED FOR CATALYTIC ACTIVITY.  
CC -!- PTM: THE N-TERMINAL IS BLOCKED.  
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE  
CC SPLICING.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19; ALSO KNOWN AS  
CC FAMILY 2 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X91349; E208113; -  
DR EMBL; U47927; G1208744; -  
DR EMBL; U47924; G1203896; -  
DR EMBL; U47924; G1203895; -  
DR EMBL; U35116; G1008542; -  
DR MIM; 601447; -  
DR PROSITE; PS00972; UCH\_2\_1; 1.  
DR PROSITE; PS00973; UCH\_2\_2; 1.  
DR PFAM; PF00442; UCH-1; 1.  
DR PFAM; PF00443; UCH-2; 1.  
DR PFAM; PF0627; UBA; 2.  
KW UBIQUITIN CONJUGATION; HYDROLASE; THIOL PROTEASE; MULTIGENE FAMILY;  
ALTERNATIVE SPLICING; ZINC.  
FT ACT\_SITE 335 335 BY SIMILARITY.  
FT ACT\_SITE 809 809 BY SIMILARITY.  
FT ACT\_SITE 818 818 BY SIMILARITY.  
FT VARSPLIC 629 652 GSLGFVGNEDSFCSPPSPPTS -> A (IN SHORT  
FT FORM).  
FT EL -> DV (IN REF. 1).  
FT CONFLICT 3 4  
FT CONFLICT 45 45 I -> V (IN REF. 1).  
FT CONFLICT 468 468 K -> R (IN REF. 3).  
FT CONFLICT 681 681 G -> D (IN REF. 3).  
SQ SEQUENCE 858 AA; 95786 MW; 1F0B83C9 CRC32;

Query Match 44.9%; Score 62; DB 1; Length 858;  
Best Local Similarity 41.2%; Pred. No. 3.84e+00;  
Matches 7; Conservative 4; Mismatches 5; Indels 1;  
Gaps 1;  
Db 47 MNTFLG-FGKQYVERHF 62  
QY 1 MERFMAEFGOGYVOTPF 17  
|:|::|||  
Search completed: Wed Sep 1 16:07:37 1999  
Job time : 11 secs.

Result No.	Score		Query Match	Length	DB	ID	Description	Pred. No.
	77	69						
1	77	55.8	860	1	059003	860AA LONG HYPOTHETICA	2.01e-02	
2	69	50.0	364	3	Q74624	MANNOSE-1-PHOSPHATE GU	5.40e-01	
3	68	49.3	178	3	060074	HYPOTHETICAL 19.4 KD P	8.03e-01	
4	66	47.8	516	10	Q42701	CYCLOHOME P450 (EC 1.	1.76e+00	
5	65	47.1	2301	10	080784	FT3P17.19 PROTEIN.	2.59e+00	
6	64	46.4	273	2	057454	HYPOTHETICAL 32.9 KD P	3.80e+00	
7	64	46.4	420	8	Q37744	PHOSPHOGLYCERATE KINAS	3.80e+00	
8	64	46.4	440	8	Q37745	PHOSPHOGLYCERATE KINAS	3.80e+00	
9	64	46.4	509	8	Q37743	PHOSPHOGLYCERATE KINAS	3.80e+00	
10	64	46.4	524	10	Q42700	CYCLOHOME P450 (EC 1.	3.80e+00	
11	63	45.7	355	5	024013	STIMILAR TO FAMILY 1 OF	5.56e+00	
12	63	45.7	455	2	Q83033	ACYLTRANSFERASE.	5.56e+00	
13	63	45.7	645	2	Q53818	PEP-FAMILY PROTEIN.	5.56e+00	
14	63	45.7	1215	2	Q48232	DNA FOR SEROTYPE B CAP	5.56e+00	
15	62	44.9	191	1	Q27457	FUCULOSE-1-PHOSPHATE	8.10e+00	
16	61	44.2	605	14	Q56954	E6, ET, E1, E2, E4, L2	1.17e+01	
17	60	43.5	205	1	Q29634	RIBONUCLEASE H II (RNHB	1.70e+01	
18	60	43.5	499	2	Q52196	ALGI.	1.70e+01	
19	60	43.5	520	2	Q51392	ALGI.	1.70e+01	
20	60	43.5	558	1	Q58315	588AA LONG HYPOTHETICA	1.70e+01	

```
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
OC HYPOCREALES; HYPOCREACEAE; HYPOCREA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RUT C-30;
RX MEDLINE: 9809839.
RA KRUSZESKA J.S., SALOHEIMO M., PENTTILA M., PALAMARCZYK G.;
RT "Isolation of a Trichoderma reesei cDNA encoding GTP:
RT alfa-d-mannose-1-phosphate guanylttransferase involved in early steps
RT of protein glycosylation.";
RL CURR. GENET. 33:445-450(1998).
DR EMBL: U89991; G3323397; -.
KW TRANSFERASE; NUCLEOTIDYLTRANSFERASE.
SQ SEQUENCE 364 AA; 40284 MW; 95D4573B CRC32;

Query Match 50.0%; Score 69; DB 3; Length 364;
Best Local Similarity 44.4%; Pred. No. 5.40e-01;
Matches 8; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

Db 60 MEKFLAEYEKYNINIEF 77
QY 1 MEKFMFEFGQGY-VQTPF 17

RESULT 3
ID O60074 PRELIMINARY; PRT; 178 AA.
AC O60074;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 19.4 KD PROTEIN.
GN SPBC13G1.12.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972H;
RA LYNE M., WOOD V., RAJANDREAM M.A., BARRELL B.G., BADCOCK K.,
RA CHURCHER C.M.;
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AL022600; E1287787; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 178 AA; 19432 MW; F002D830 CRC32;

Query Match 49.3%; Score 68; DB 3; Length 178;
Best Local Similarity 47.1%; Pred. No. 8.03e-01;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 93 MDKFEAQFDDVNYQTGY 109
QY 1 MEKFMFEFGQGY-VQTPF 17

RESULT 4
ID Q42701 PRELIMINARY; PRT; 516 AA.
AC Q42701;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 (EC 1.14.14.1) (FRAGMENT).
GN CYP72C.
OS CATHARANTHUS ROSEUS (ROSY PERIWINKLE) (MADAGASCAR PERIWINKLE).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; GENTIANACEAE; GENTIANACEAE; APOCYNACEAE; CATHARANTHUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CP3;
RA MANGOLD U., EICHEL J., BATSCHAUER A., LANZ T., KAISER T.,
RA SPANGENBERG G., WERCK-REICHART D., SCHROEDER J.;
RL PLANT SCI. 96:129-136(1994).
```

```
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: L19075; G404690; -.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
DR PFAM: PF00067; P450; 1.
DR MENDEL: 11605; CATO:1113;14.
KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME.
FT NON_TER 1
FT BINDING 459 459 HEME (BY SIMILARITY).
SQ SEQUENCE 516 AA; 59720 MW; AD63FB23 CRC32;

Query Match 47.8%; Score 66; DB 10; Length 516;
Best Local Similarity 29.4%; Pred. No. 1.76e+00;
Matches 5; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

Db 477 LORFKFDVAPSYVHAPE 493
QY 1 MEKFMFEFGQGY-VQTPF 17

RESULT 5
ID O80784 PRELIMINARY; PRT; 2301 AA.
AC O80784;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE F13P17.19 PROTEIN.
GN F13P17.19.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SIKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RT "Arabidopsis thaliana chromosome II BAC F13P17 genomic sequence.";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AC004481; G3337366; -.
SQ SEQUENCE 2301 AA; 253398 MW; FB931E6C CRC32;

Query Match 47.1%; Score 65; DB 10; Length 2301;
Best Local Similarity 46.7%; Pred. No. 2.59e+00;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 977 NFLGRFGAGYVSDTF 991
QY 3 KFMFEFGQGY-VQTPF 17

RESULT 6
ID O67454 PRELIMINARY; PRT; 273 AA.
AC O67454;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 32.9 KD PROTEIN.
GN AQ1477.
OS AQUIFEX AEOLICUS.
OC BACTERIA; AQUIFICALES; AQUIFICACEAE; AQUIFEX.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RX MEDLINE: 98196666.
RA DECKERT G.E., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL NATURE 392:353-358(1998).
RN [2]
RP SEQUENCE FROM N.A.
```



```

Db 264 KFNLFSHTVQOF 278
QY 3 KFMAEFGQGVQTFP 17

RESULT 12
ID 084013 PRELIMINARY; PRT: 455 AA.
AC 084013;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ACYLTRANSFERASE.
GN HTRB.
OS CHLAMYDIA TRACHOMATIS.
OC BACTERIA; CHLAMYDIALES; CHLAMYDIACEAE; CHLAMYDIA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=D/UN-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATSOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT Chlamydia trachomatis.";
RL SCIENCE 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UN-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATSOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; AE001275; G3328398; -.
KW TRANSFERASE; ACYLTRANSFERASE.
SQ SEQUENCE 455 AA; 52058 MW; 18AB7CEE CRC32;

Query Match 45.7%; Score 63; DB 2; Length 455;
Best Local Similarity 29.4%; Pred. No. 5,56e+00;
Matches 5; Conservative 9; Mismatches 3; Indels 0; Gaps

Db 343 LORFLTEFGFYDASL 359
QY 1 MEKPMAEFGQGVQTFP 17

RESULT 13
ID 053818 PRELIMINARY; PRT: 645 AA.
AC 053818;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PPE-FAMILY PROTEIN.
GN MY041.29.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIACEAE;
OC ACTINOMYCETALES; CORYNEBACTERIACEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA HAWLIN N., CHURCHER C.M.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA COLE S.T., PARKHILL J., BARRELL B.G., RAJANDREAN M.A.;
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE; 96181548.
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA COLE S.T.;
RT "An integrated map of the genome of the tubercle bacillus.

```

RT Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium  
 leprae.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).  
 DR EMBL; AL021958; E1253293; -;  
 SQ SEQUENCE 645 AA; 62693 MW; 04F5F7F1 CRC32;

Query Match 45.7%; Score 63; DB 2; Length 645;  
 Best Local Similarity 46.7%; Pred. No. 5.56e+00;  
 Matches 7; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

Db 592 FMGFGNTGFLOSFG 606

Qy 4 FMAEFG-QYVQTFP 17

#### RESULT 14

ID O48232 PRELIMINARY; PRT; 1215 AA.

AC O48232;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE DNA FOR SEROTYPE B CAPSULATION LOCUS.

OS HAEMOPHILUS INFLUENZAE.

OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURACEAE;

OC HAEMOPHILUS.

RN [1]

RC SEQUENCE FROM N.A.

RP STRAIN-RM135;

RX MEDLINE: 95272382.

RA VAN ELDRE J., BROPHY L., LOYNDS B., CELIS P., KROLL J.S., MOXON E.R.,

RA HANCOCK I., CARMAN S.;

RT "Region II of the Haemophilus influenzae type be capsulation locus is

involved in serotype-specific polysaccharide synthesis.";

RL MOL. MICROBIOL. 15:107-118(1995).

DR EMBL: X78559; G471236; -;

SQ SEQUENCE 1215 AA; 142323 MW; DF50121E CRC32;

Query Match

45.7%; Score 63; DB 2; Length 1215;

Best Local Similarity 52.6%; Pred. No. 5.56e+00;

Matches 10; Conservative 2; Mismatches 5; Indels 2; Gaps 2;

Db 811 KEKFAFVEFGVGTDTF 829

Qy 1 MEKF-MAEF-QGVVQTFP 17

#### RESULT 15

ID O27457 PRELIMINARY; PRT; 191 AA.

AC O27457;

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE FUCULOSE-1-PHOSPHATE ALDOLASE.

GN MTH1406.

OS METHANOBACTERIUM THERMOAUTOTROPHICUM.

OC ARCHAEA; EURYARCHAEOTA; METHANOBACTERIALES; METHANOBACTERIACEAE;

OC METHANOBACTERIUM.

RN [1]

RC SEQUENCE FROM N.A.

RP STRAIN-DELTA H;

RX MEDLINE: 98037514.

RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,

RA ALDREDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,

RA HARRISON D., HOANG L., KEAGLE P., LUMM W., POTHIER B., QIU D.,

RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,

RA JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,

RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,

RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;

RT "Complete genome sequence of Methanobacterium thermoautotrophicum

deltah: functional analysis and comparative genomics.";

RL J. BACTERIOL. 179:7135-7155(1997).

DR EMBL; AE000903; G2622518; -;

DR PFAM; PF00596; Aldolase-II; 1.

SQ SEQUENCE 191 AA; 20702 MW; BC8B78DC CRC32;

Query Match 44.9%; Score 62; DB 1; Length 191;  
 Best Local Similarity 44.4%; Pred. No. 8.10e+00;  
 Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 2;

Db 111 MEGFMG-LGRGYIPMPY 127

Qy 1 MEKFMABFGQYVQ-TPF 17

Search completed: Wed Sep 1 16:08:14 1999  
 Job time : 21 secs.

\*\*\*\*\*

W P E R E H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:11:25 1999; MasPar time 5.59 Seconds

Tabular output not generated. 213.200 Million cell updates/sec

Title: >PCT-US99-13024-2

Description: (1-56) from PCTUS9913024.pep (4 of 12)

Perfect Score: 402

Sequence: 1 MEKFMARFGQGVQTPFLSE.....STAGPSYVKFQDNPVGSQTF 56

Scoring table: PAM 150

Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 25.232; Variance 100.718; scale 0.251

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	300	74.6	65.19	R97371	Phage T4 ORF gene pr	1.93e-20
2	81	20.1	3011	12	Hepatitis C virus gen	1.51e+01
3	81	20.1	3011	8	HCV genomic amino aci	1.51e+01
4	80	19.9	170	1	Sequence of hepatitis	1.82e+01
5	80	19.9	170	1	Sequence encoded in t	1.82e+01
6	80	19.9	411	16	HCV NS5 domain antige	1.82e+01
7	80	19.9	499	38	HCV NS5 non-structural pr	1.82e+01
8	80	19.9	516	6	HCV NS5 encoded by ph	1.82e+01
9	80	19.9	516	7	HCV CKS-NS5E fusion p	1.82e+01
10	80	19.9	516	6	HCV CKS-NS5E recombin	1.82e+01
11	80	19.9	516	6	HCV CKS-NS5E recombin	1.82e+01
12	80	19.9	516	6	HCV CKS-NS5E recombin	1.82e+01
13	80	19.9	798	7	HCV CKS-NS5 EF fusion	1.82e+01
14	80	19.9	1766	1	Sequence encoded in t	1.82e+01
15	80	19.9	1786	1	Protein sequence of h	1.82e+01
16	80	19.9	2261	1	Peptide encoded by co	1.82e+01

17	80	19.9	2301	1	P92047	Sequence encoded in t
18	80	19.9	2435	5	R25135	HCV polypeptide 1.
19	80	19.9	2436	1	P92050	Sequence encoded in t
20	80	19.9	2436	5	R28582	HCV amino acid sequen
21	80	19.9	2462	1	P90288	Peptide encoded by co
22	80	19.9	2772	2	R08123	Hepatitis C virus pol
23	80	19.9	2816	7	R34009	HCV-1 polypeptide.
24	80	19.9	2894	5	R24440	Composite hepatitis C
25	80	19.9	2894	13	R70230	Composite hepatitis C
26	80	19.9	2955	2	R08124	Hepatitis C virus put
27	80	19.9	3011	26	W34480	HCV polypeptide.
28	80	19.9	3011	4	R21519	Compiled HCV sequence
29	80	19.9	3011	28	W40038	HCV polypeptide.
30	80	19.9	3011	16	R09031	Hepatitis C virus pol
31	80	19.9	3011	6	R31621	Hepatitis C virus (HC
32	79	19.7	458	11	R56975	Variant vitamin D bin
33	79	19.7	458	4	R22278	Human gp.-specific co
34	79	19.7	458	23	W10561	Vitamin D3-binding pr
35	79	19.7	458	11	R56976	Variant vitamin D bin
36	79	19.7	458	4	R22279	Human gp.-specific co
37	79	19.7	1464	10	R55529	Human NMDA R2A recept
38	78	19.4	3011	36	W77397	Hepatitis C virus H77
39	78	19.4	3011	4	R22154	NANBV Hutch c59 isola
40	78	19.4	3011	8	R40119	HCV genomic amino aci
41	78	19.4	3011	14	R79232	HCV sequence.
42	78	19.4	3011	36	W77398	Hepatitis C virus-H C
43	77	19.2	135	35	W76234	Bacterial periplasmic
44	77	19.2	1464	18	R80970	Human excitatory amin
45	77	19.2	1464	38	W85576	Human N-methyl-D-aspa

# ALIGNMENTS

RESULT 1  
ID R97371 standard; Protein; 65 AA.  
AC R97371;  
DE 07-JAN-1997 (first entry)  
DT Phage T4 ORF gene product. gp34.  
KW filter; tail fibre protein; nanotechnology; nano-structure;  
OS Bacteriophage T4.  
OS Bacteriophage T4.  
PF 13-OCT-1995; U13023.  
PR 13-OCT-1994; US-322760.  
PA (GOLD/) GOLDBERG E B.  
PI GOLDBERG EB;  
DR WPI; 96-221942/22.  
DR N-PSDB; T29053.  
PT New proteins derived from T4 phage tail fibre proteins - that can self assemble into nano-structure(s), useful as filters etc, also corresponding DNA  
PS Claim 7; Fig 7; 83pp; English.  
CC A protein (R97371) of unspecified function is the product of open reading frame x of the tail fibre protein gene region (see also T29053) of phage T4. This gene region also includes open reading frames for tail fibre proteins (see also R97370 and R97372-74).  
CC Tail fibre proteins (native or modified) can be produced in large quantities in microbial cells and used as bulding blocks of strong, stable nanostructures.  
SQ Sequence 65 AA;

Query Match 74.6%; Score 300; DB 19; Length 65;  
Best Local Similarity 82.1%; Pred. No. 1.93e-20;

Matches 46; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Db 1 mekfmairtricpnaalisesnrykisiagscplstagspyvxfqndpvgstff 56

QY 1 MEKFMARFGQGVQTPFLSESNRYKISAGSCPLSTAGSPYVKFQDNPVGSQTF 56

RESULT 2

ID R66995 standard; Protein; 3011 AA.



[illegible]

```

RESULT      8
ID R33441 standard; protein; 516 AA.
AC R33441:
DT 30-JUN-1993 (first entry)
DE HCV NS5 encoded by pHCV45 clone.
KW Monoclonal antibody; hybridoma; E.coli; diagnosis.
OS Hepatitis C virus.
PN W09304084-A.
PD 04-MAR-1993.
PF 21-AUG-1992; U06965..
PR 21-AUG-1991; US-748563.
PA (ABBO ) ABBOTT LAB.
PI Dailey SH, Desai SM, Devare SG, Johnson JE, Mehta SU;
DR WPT: 93-093937/11.
PT Monoclonal antibodies to hepatitis C virus NS5 proteins - used in
PT differentiating hepatitis C infections and differentiation studies
PS Disclosure; Page 27; 38pp; English.
CC The sequence is that of an E. coli derived fragment of Hepatitis C
CC virus NS5. The sequence may be used to raise monoclonal antibodies
CC which specifically bind HCV NS5 antigen and to produce the hybridoma
CC cell lines ATCC HB 10854 and HB 10855.
CC See also R33442-4.
SQ Sequence 516 AA;

Query Match          19.9%; Score 80; DB 6; Length 516;
Best Local Similarity 37.8%; Pred.No.1.82e+01;
Matches 17; Conservative 11; Mismatches 13; Indels 4; Gaps 4

Db 433 vpspeffteidgrvrlh-rfapppckpllreevsfrvgihyepvqsg 476
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 13 VQTP-FLSESNSVRYKTSIAGSC-PLSTAGPSY-VKFQDNPVCSQ 54

```

RESULT 9  
ID R33633 standard; Protein; 516 AA.  
AC R33633.

DE	HCV CKS-NS5E fusion protein pHCV45.
KE	Hepatitis C virus; NANBH; non-A, non-B hepatitis; CMP-KDO synthetase;
KW	non-structural protein; pHCV-45; diagnosis; CKS fusion protein;
KW	CYP: CMP-3-deoxy-manno-octulosonate cytidyl transferase;
KW	Immunocassay; ss.
OS	Hepatitis C Virus.
PN	WO9304089-A.
PD	04-MAR-1993.
PF	21-AUG-1992; U06964.
PR	21-AUG-1991; US-748565.
PA	(ABBO ) ABBOTT LAB.
PI	Dailey SH, Desal SM, Devare SG;
DR	WPI: 93-093942/11.
DR	N-PSDB; Q38267.
PT	New recombinant NS5 region antigens - for hepatitis C assay for
PT	detecting hepatitis C virus infections
PS	Example 7; Page 56-57; 164pp; English.
CC	Eight oligonucleotides representing amino acids 1932-2191 of the
CC	HCV genome were ligated together and cloned as a 739bp EcoRI/BamHI
CC	fragment into the CKS fusion vector pJO200. The complete DNA
CC	sequence of this plasmid is designated pHCV-45. The resultant
CC	fusion protein HCV CKS-NS5E consists of 239 amino acids of CKS,
CC	nine amino acids contributed by linker DNA sequences and 260 amino
CC	acids from the HCV NS4/NS5 region (amino acids 1932-2191). The
CC	recombinant HCV CKS-NS5E antigen was immunoreactive when screened

SQ Sequence 516 AA;   
 Query Match 19.9%; Score 80; DB 7; Length 516;   
 Best Local Similarity 37.98; Pred. No. 1.82e+01;   
 Matches 17; Conservative 13; Indels 4; Gaps 4;   
 Db 433 vpspefftdgvrhl-rfappckllreevsfrvlgheypvgsg 476

```
CC HCV.
SQ Sequence      516 AA;

Query Match          19.9%; Score 80; DB 6; Length 516;
Best Local Similarity 37.8%; Pred. No. 1.82e+01;
Matches              17; Conservative 11; Mismatches 13; Indels 4; Gaps 4;

Db    433 vpspeffteidgvrhl-rfappckpllireevsfvrglhheypvgsq 476
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY    13 VQTPLFLESNSVRYKISAGSC-PLSTAGPSY-VKFQDNPVGQSQ 54

RESULT 12
ID     R21566 standard; Protein; 516 AA.
AC     R21566;
DT     09-JUN-1992 (first entry)
DE     HCV CKS-NS5E - pHCV-45.
KW     Hepatitis C virus; antigen; diagnosis; inhibitor; CMP-KDO synthase;
KW     CKS; HCV CKS-NS5E; NANBH.
FH     Key Location/Qualifiers
FT     peptide 1..239
FT           /label= CKS
FT     region 240..248
FT           /label= linker
FT           /note= "see CC"
FT     peptide 249..508
FT           /label= NS5-E
FT           /note= "HCV region NS4/NS5, amino acids 1932-2191"
FT     region 509..516
FT           /label= linker
FT           /note= "see CC"
EP-472207-A.
PD     26-FEB-1992.
PN     23-AUG-1991; 114161.
PF     24-AUG-1990; US-572822.
PR     07-NOV-1990; US-614069.
PA     (Devare SG, Desai SM, Casey JM, Dawson GJ, Lesniewski RR;
PI     Dailey SH, Gutierrez RA, Stewart JL;
PI     WPI: 92-066430/09.
DR     N-PSDB: Q21679.
PT     Recombinant hepatitis C virus antigens - produced as fusion
PT     proteins and representing distinct antigenic regions of the HCV
PT     genome
PS     Disclosure; Fig 26, Page 61-63; 115pp; English.
CC     Note: according to the diagrammatic illustration of this sequence,
CC     both linkers comprise 9 amino acids.
CC     The polypeptide (mol.wt. 55 kD) represents a distinct antigenic
CC     region of the HCV genome and can be used for the detection of
CC     antibodies and antigens for early diagnosis of HCV infection.
CC     The polypeptide can also be used to develop specific inhibitors of
CC     viral replication and for therapeutic purposes.
SQ     Sequence 516 AA;

Query Match          19.98; Score 80; DB 4; Length 516;
Best Local Similarity 37.8%; Pred. No. 1.82e+01;
Matches              17; Conservative 11; Mismatches 13; Indels 4; Gaps 4;

Db    433 vpspeffteidgvrhl-rfappckpllireevsfvrglhheypvgsq 476
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY    13 VQTPLFLESNSVRYKISAGSC-PLSTAGPSY-VKFQDNPVGQSQ 54

RESULT 13
ID     R33630 standard; protein; 798 AA.
AC     R33630;
DT     13-JUL-1993 (first entry)
DE     HCV CKS-NS5 EF fusion protein pHCV59.
KW     Hepatitis C virus; NANBH; non-A, non-B hepatitis; CMP-KDO synthetase;
KW     non-structural protein; pHCV-59; diagnosis; immunoassay;
KW     CTP;CMP-3-deoxy-manno-octulosonate cytidyl transferase.
OS     Hepatitis C Virus.
SN     WO9304089-A.
PN
```

```

04-MAR-1993.
PD 21-AUG-1992; U06964.
PF PR 21-AUG-1991; US-748565.
PA (ABBO ) ABBOTT LAB.
PI Dailey SH, Desal SM, Devare SG;
DR WPI; 93-093942/11.
PT New recombinant NS5 region antigens - for hepatitis C assay for
PT detecting hepatitis C virus infections
PS Claim 1; Page 36-38; 164pp; English.
CC Eight oligonucleotides representing amino acids 2188-2481 of the HCV
CC genome were ligated together and cloned as a 895bp EcoRI-BamHI
CC fragment in the vector pJO200 (i.e. pBR322 having a modified lac
CC promoter fused to part of the CKS gene and contg. a synthetic linker)
CC to produce pHCV-48. The construct pHCV-45 was prepared by cloning a
CC 793bp fragment representing HCV amino acids 1932-2191 into pJO200.
CC The SalI-BamHI fragment from pHCV-48 was ligated into the SalI-BamHI
CC vector backbone of pHCV-45 to give an insert representing AAs 1932-
CC 2491 of HCV. The resulting 1650bp EcoRI-BamHI fragment was cloned
CC into pJO200 to produce pHCV-59. The amino acid sequence of the
CC resultant fusion antigen HCV CKS-NS5 EF consists of 239 amino acids
CC of CKS, 9 linker amino acids and 550 amino acids from the NS5
CC region of the HCV genome. The fusion protein is useful in
CC immunoassays for detecting anti-HCV antibodies.
SQ Sequence 798 AA;

Query Match          19.9%; Score 80; DB 7; Length 798;
Best Local Similarity 37.8%; Pred. No. 1.82e+01;
Matches      17; Conservative    11; Mismatches   13; Indels     4; Gaps     4;

Db       433 vpspeffteidgrlh-rfappckpllreesfrvrlghieyvpqsg 476
        |:|::||:::|| |::|| |::|| |::|| |::|| |::|| |
QY      13 VQTTP-FLSESNVRYKTSIAGSC-PLSTAGPSY-VKFQDNPVGSG 54

Search completed: Wed Sep 1 16:12:13 1999
Job time : 48 secs.
```

```

RESULT      14
ID          P92041 standard; protein; 1766 AA
AC          P92041
DE          02-MAR-1990 (first entry)
DT          DE
TE          Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones
KW          Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH).
OS          Hepatitis C virus.
PN          EP-318216-A.
PD          31-MAY-1989.
PF          18-NOV-1988: 310922.
PR          14-NOV-1988: US-271450; US-122714.
PA          (CHIR) Chiron Corp.
PI          Houghton M, Choo q-L, Kuo G;
DR          WPI; 89-159274/22.
DR          N-PSTB; N92097
PT          Purified hepatitis C virus
CC          - and associated nucleic acids and polypeptide(s)
PS          Clam13; Figure 26-1, 26-2, 26-3, 26-4, 26-5, 26-6; 139pp; English.
CC          14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f,
CC          33f, 33g and 39c.
CC          It is the sequence encoded in the open reading frame of hepatitis C virus
CC          cDNA inserts in clones 14i,m 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36,
CC          8l, 32, 33b, 25c, 14c, 8f, a33f, 33g and 39c. It is antigenic and could
CC          be used in immunoassay reagents and vaccines and to generate antibodies
CC          useful in diagnosis and passive immunotherapy for HCV infection/non-A,
CC          non-B hepatitis.
SQ          Query       1766 AA;
SC          Sequence    1766 AA;

Query Match           19.9%; Score 80; DB 1; Length 1766;
Best Local Similarity 37.8%; Pred. No. 1.82e+01;
Matches              17; Conservative     11; Mismatches 13; Indels   4; Gaps  4;

Db      1380 vspsefftdgvrlrh-vfappckpllrveevsfvrglhvpvgsg 1423
               | : : : | : ||| | : | ||| | : | ||| |
Oy      13 VQTP-FLSESNVRKYSIAGSC-PLSTAGPSPY-VKFQDNFVGSG 54

```

RESULT 15  
ID P90158 standard; protein; 1786 AA.



```
##status      preliminary
##molecule_type DNA
##residues    1-869 #label JOR
##cross-references GB:M55661; NID:g145507; PID:g145510
##experimental_source enterotoxigenic strain, CrAY/I-ST plasmid NTP113
##note        sequence extracted from NCBI backbone (NCBIN:108960,
               NCBIP:108971)

GENETICS
#gene         cfaC
#genome       plasmid
SUMMARY
#length 869 #molecular-weight 97830 #checksum 9755
Query Match      21.1%; Score 85; DB 2; Length 869;
Best Local Similarity 32.4%; Pred. No. 2.31e-01;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;

Db 158 AFTQSQTINLSDGKYKRLISGNSALG1TDTSY 191
      ::::: : : : : : : : : : : : : : : : : :
QY 11 GYVOTPFLSENSVRYK-ISIAGSCLSTAGPSY 43

RESULT 3 ~
ENTRY
TITLE      I38375 #type fragment
ALTERNATE_NAMES tyrosine kinase - human (fragment)
ORGANISM      #formal_name Homo sapiens #common_name man
DATE          29-May-1998 #sequence_revision 29-May-1998 #text_change
               10-Jul-1998
ACCESSIONS    I38375
REFERENCE     I38372
#authors      Haire, R.N.; Ohta, Y.; Lewis, J.E.; Fu, S.M.; Kroisel, P.;
               Litman, G.W.
#journal      Hum. Mol. Genet. (1994) 3:897-901
#title        TXK, a novel human tyrosine kinase expressed in T cells
               shares sequence identity with Tec family kinases and maps
               to 4p12
#cross-references MUID:95038742
#accession     I38375
##status      preliminary; translated from GB/EMBL/DBDJ
##molecule_type DNA
##residues    1-81 #label RES
##cross-references EMBL:U07794; NID:g508219; PID:g508224

GENETICS
#gene         GDB:TXK
#map_position 4p12-4p12
#introns      25/1
CLASSIFICATION #superfamily protein-tyrosine kinase tec; pleckstrin repeat
               homology; protein kinase homology; SH2 homology; SH3
               homology
FEATURE
32-81
SUMMARY      #domain protein kinase homology (fragment) #label KIN
               #length 81 #checksum 7751
Query Match      20.4%; Score 82; DB 2; Length 81;
Best Local Similarity 44.4%; Pred. No. 6.47e-01;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 5 LRPVGLMGSLCPATAGFSYEKWEIDP 31
      : : : : : : : : : : : : : : : :
QY 24 VRYKISAGSCLSTAGPSVVKFQDNP 50

RESULT 4
ENTRY
TITLE      I84483 #type complete
ALTERNATE_NAMES tyrosine kinase - human
ORGANISM      #formal_name Homo sapiens #common_name man
DATE          29-May-1998 #sequence_revision 29-May-1998 #text_change
               12-Feb-1999
ACCESSIONS    I84483
REFERENCE     I38372
#authors      Haire, R.N.; Ohta, Y.; Lewis, J.E.; Fu, S.M.; Kroisel, P.;
               Litman, G.W.
#journal      Hum. Mol. Genet. (1994) 3:897-901

TXK, a novel human tyrosine kinase expressed in T cells
shares sequence identity with Tec family kinases and maps
to 4p12
#cross-references MUID:95038742
#accession     I84483
##status      preliminary; translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues    1-527 #label RES
##cross-references GB:L35268; NID:g623442; PID:g623443
CLASSIFICATION #superfamily protein-tyrosine kinase tec; pleckstrin repeat
               homology; protein kinase homology; SH2 homology; SH3
               homology
KEYWORDS
FEATURE
89-137
150-246
269-527
277-285
SUMMARY      #domain SH3 homology #label SH3\
               #domain SH2 homology #label SH2\
               #domain protein kinase homology #label KIN\
               #region protein kinase ATP-binding motif
               #length 527 #molecular-weight 61108 #checksum 785

Query Match      20.1%; Score 81; DB 2; Length 527;
Best Local Similarity 44.4%; Pred. No. 9.06e-01;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 242 LRPVGLMGSLCPATAGFSYEKWEIDP 268
      : : : : : : : : : : : : : : : :
QY 24 VRYKISAGSCLSTAGPSVVKFQDNP 50

RESULT 6
ENTRY
TITLE      I49133 #type complete
```

QY	13	VQTP-FLSESNVRYKISAGSC-PLSTAGPSY-VKFDNPVGSQ	54
RESULT	8		
ENTRY			
TITLE			
ALTERNATE_NAMES			
ORGANISM			
DATE			
ACCESSIONS			
REFERENCE			
authors			
submission			
description			
accession			
residues			
cross-references			
experimental_source			
REFERENCE			
authors			
journal			
title			
cross-references			
accession			
status			
molecule_type			
residues			
cross-references			
experimental_source			
REFERENCE			
authors			
journal			
title			
cross-references			
accession			
status			
molecule_type			
residues			
cross-references			
GENETICS			
gene			
map_position			
note			
KEYWORDS			
FEATURE			
331-396			
SUMMARY			
Query Match			
Best Local Similarity			
Matches			
Db	622	QGYSGQPIPNQSYG-RYQTSIPPPPOQQIPQSGYGRYQAGP	663
QY	10	QGY-VQTPF-LSESNVRYKISAGSCPLSTAGPSYVKFDNP	50
RESULT	9		
ENTRY			
TITLE			
CONTAINS			
ORGANISM			
DATE			

```

22-May-1998
A39166; PQ0403; PQ0404
REFERENCE
#authors
Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.;
Dong, C.; Gallegos, C.; Colt, D.; Medina-Selby, A.; Barr,
P.J.; Weiner, A.J.; Bradley, D.W.; Kuo, G.; Houghton, M.
Proc. Natl. Acad. Sci. U.S.A. (1991) 88:2451-2455
#journal
Genetic organization and diversity of the hepatitis C virus.
#title
Cross-references MUID:91172826
#accession
A39166
##molecule_type mRNA
##residues 1-3011 #label CHO
##cross-references GB:M62321; NID:g329873; PID:g329874
PQ0393
REFERENCE
#authors
Chan, S.W.; McMahon, F.; Holmes, E.C.; Dow, B.; Peutherer,
J.F.; Follett, E.; Yap, P.L.; Simmonds, P.
J. Gen. Virol. (1992) 73:1131-1141
#journal
Analysis of a new hepatitis C virus type and its phylogenetic
relationship to existing variants.
#title
PQ0403
#accession
PQ0403
##molecule_type genomic RNA
##residues 1577-1633 #label CHA
##cross-references DBJ:D10128
##experimental_source isolates E-b16
#accession
PQ0404
##status preliminary
##molecule_type genomic RNA
##residues 1577-1633 #label CH2
##experimental_source isolates E-b17
CLASSIFICATION
#superfamily hepatitis C virus genome polyprotein; DEAD/H box
helicase homology
KEYWORDS
ATP; capsid protein; envelope protein; glycoprotein;
nonstructural protein; p-loop; polyprotein; transmembrane
protein
FEATURE
1-115      *product capsid protein C #status predicted #label CPC\
116-191    *product envelope protein M #status predicted #label
EPW\
192-389    *product major envelope protein E #status predicted
          #label MEE\
390-729    *product nonstructural protein NS1 #status predicted
          #label NS1\
730-1006   *product nonstructural protein NS2 #status predicted
          #label NS2\
1007-1615  *product nonstructural protein NS3 #status predicted
          #label NS3\
1230-1493  *domain DEAD/H box helicase homology #label DEAD\
1230-1237  *region nucleotide-binding motif A (#loop)\
1312-1317  *region nucleotide-binding motif B\
1316-1319  *region DEHX motif\
1616-1862  *product nonstructural protein NS4a #status predicted
          #label N4A\
1863-2013  *product nonstructural protein NS4b #status predicted
          #label N4B\
          *product nonstructural protein NS5 #status predicted
          #label NS5\
2014-3011
196,209,234,305,
325,417,423,430,
448,476,532,540,
556,576,623,645,
1213,1255,2041,
2077,2240,2364,
2550,2789
#binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY      #length 3011 #molecular-weight 327199 #checksum 8610
Query Match 19.9%; Score 80; DB 1; Length 3011;
Best Local Similarity 37.8%; Pred. No. 1.27e+00;
Matches 17; Conservative 11; Mismatches 13; Indels 4; Gaps 4;
Db 2116 VPSPEFFTELDGRLH-RFAPCKPLLRVSVFRGLHPYGVQS 2159
| : | : | : ||||| : | : | : | : | : | : | : | : |
QY 13 VQTP-FLESNSVRYSKIAGSC-PLSTAGPSY-VKFDNPVGVS 54

```

[illegible]







OY 1 MEKFMAEFGGYYVOTPFLESNSVRYKISIAAGSCPLS 37

Search completed: Wed Sep 1 16:11:07 1999  
Job time : 16 secs.

\*\*\*\*\*  
WILEY  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:09:43 1999; MasPar time 3.74 Seconds  
Tabular output not generated. 422.797 Million cell updates/sec

Title: >PCT-US99-13024-2  
Description: (1-56) from PCTUS9913024.pep (4 of 12)  
Perfect Score: 402  
Sequence: 1 MEKFAEFGGQYVQTPFLSE.....STAGPSYVKFDNPNVGSQTF 56

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 36.535; Variance 57.714; scale 0.633

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result	No.	Score	Query %	Match	Length	ID	Pred. No.
1	300	74.6	65	1	Y15A_BPT4	HYPOTHETICAL 7.3 KD PR	4.61e-49
2	85	21.1	869	1	CFAC_ECOLI	CFA/I FIMBRIAL SUBUNIT	6.62e-02
3	82	20.4	527	1	TXK_HUMAN	TYROSINE-PROTEIN KINAS	2.04e-01
4	81	20.1	527	1	TXK_MOUSE	TYROSINE-PROTEIN KINAS	2.96e-01
5	80	19.9	802	1	NAB3_YEAST	NUCLEAR POLYADENYLATED	4.27e-01
6	80	19.9	3011	1	POLG_HCV1	GENOME POLYPROTEIN [CO	4.27e-01
7	79	19.7	472	1	VTDB_MOUSE	VITAMIN D-BINDING PROT	6.14e-01
8	79	19.7	474	1	VTDB_HUMAN	VITAMIN D-BINDING PROT	6.14e-01
9	78	19.4	353	1	YQHT_BACSU	PUTATIVE PEPTIDASE IN	8.81e-01
10	78	19.4	476	1	VTDB_RABIT	VITAMIN D-BINDING PROT	8.81e-01
11	78	19.4	3011	1	POLG_HCVH	GENOME POLYPROTEIN [CO	8.81e-01
12	77	19.2	1464	1	NMEL_MOUSE	GLUTAMATE [NMDA] RECEP	1.26e+00
13	77	19.2	1464	1	NMEL_RAT	GLUTAMATE [NMDA] RECEP	1.26e+00
14	76	18.9	338	1	TCPF_VIBCH	TOXIN CORRELATED PIL	1.79e+00
15	75	18.7	575	1	HENA_INBBO	HEMAGGLUTININ PRECURSO	2.54e+00
16	75	18.7	591	1	EYAL_MOUSE	EYES ABSENT HOMOLOG 1.	2.54e+00
17	75	18.7	592	1	EYAL_HUMAN	EYES ABSENT HOMOLOG 1.	2.54e+00
18	75	18.7	984	1	DPOL_NPVAC	DNA POLYMERASE (EC 2.7	2.54e+00
19	75	18.7	986	1	DPOL_NPVBM	DNA POLYMERASE (EC 2.7	2.54e+00
20	74	18.4	171	1	YC82_PSEDE	HYPOTHETICAL 19.0 KD P	3.60e+00
21	73	18.2	105	1	CUI2_HVACE	FLEXIBLE CUTICLE PROTE	5.07e+00
22	73	18.2	130	1	RS8_METVA	30S RIBOSOMAL PROTEIN	5.07e+00
23	73	18.2	339	1	PHR_BACFI	DEOXYRIBODIPYRIMIDINE	5.07e+00

24	73	18.2	706	1	CATE_MYCAV	CATALASE HP11 (EC 1.11	5.07e+00
25	73	18.2	888	1	RIR1_YEAST	RIBONUCLEOSIDE-DIPHOSP	5.07e+00
26	73	18.2	1363	1	ILPR_BRALA	INSULIN-LIKE PEPTIDE R	5.07e+00
27	73	18.2	1693	1	POLN_HEVMY	NON-STRUCTURAL POLIPRO	5.07e+00
28	72	17.9	260	1	COAT_MDV	COAT PROTEIN.	7.11e+00
29	72	17.9	282	1	Y265_MYCPN	HYPOTHETICAL PROTEIN M	7.11e+00
30	72	17.9	345	1	GBA4_DICDI	GUANINE NUCLEOTIDE-BIN	7.11e+00
31	72	17.9	345	1	HEMA_INBFU	HEMAGGLUTININ (FRAGMEN	7.11e+00
32	72	17.9	347	1	HEMA_INBF6	HEMAGGLUTININ PRECURSO	7.11e+00
33	72	17.9	347	1	HEMA_INBF9	HEMAGGLUTININ PRECURSO	7.11e+00
34	72	17.9	354	1	HEM6_MOUSE	COPROPORPHYRINOGEN III	7.11e+00
35	72	17.9	457	1	EMB8_PICGL	LATE EMBRYOGENESIS ABU	7.11e+00
36	72	17.9	476	1	VTDB_RAT	VITAMIN D-BINDING PROT	7.11e+00
37	72	17.9	631	1	TEC_HUMAN	TYROSINE-PROTEIN KINAS	7.11e+00
38	72	17.9	695	1	GFAL_SCHPO	PUTATIVE GLUCOSAMINE--	9.95e+00
39	71	17.7	315	1	TRXB_MYCPN	THIOREDOXIN REDUCTASE	9.95e+00
40	71	17.7	318	1	NULM_HORSE	NADH-UBIQUINONE OXIDOR	9.95e+00
41	71	17.7	347	1	HEMA_INBF3	HEMAGGLUTININ PRECURSO	9.95e+00
42	71	17.7	588	1	CAR7_CANAL	CANDIDAPEPSIN 7 PRECUR	9.95e+00
43	71	17.7	611	1	PES4_YEAST	PES4 PROTEIN (DNA POLY	9.95e+00
44	71	17.7	935	1	YJ31_YEAST	HYPOTHETICAL 108.4 KD	9.95e+00
45	71	17.7	3412	1	POLG_TBEVS	GENOME POLYPROTEIN [CO	9.95e+00

ALIGNMENTS

RESULT 1	STANDARD;	PRT;	65 AA.
ID Y15A_BPT4			
AC P39509;			
DT 01-FEB-1995 (REL. 31, CREATED)			
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)			
DE HYPOTHETICAL 7.3 KD PROTEIN IN GP34-GP35 INTERGENIC REGION.			
GN Y15A OR 34.1.			
OS BACTERIOPHAGE T4.			
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE;			
OC T4-LIKE PHAGES.			
RN [1]			
RP SEQUENCE FROM N.A.			
RA KUTTER E.M.			
RL SUBMITTED (NOV-1994) TO THE SWISS-PROT DATA BANK.			
KW HYPOTHETICAL PROTEIN.			
SQ SEQUENCE 65 AA; 7334 MW; C2D7CE2D CRC32;			
Query Match	74.6%;	Score 300;	DB 1; Length 65;
Best Local Similarity	82.1%;	Pred. No. 4.61e-49;	
Matches	46;	Conservative 3;	Mismatches 7; Indels 0; Gaps 0;
Db 1 MEKFAEIMTRICPNAILSESNVRYKISAGSCPLSTAGSPSYVKFDNPNVGSQTF 56			
QY 1 MEKFAEFGGQYVQTPFLSESNVRYKISAGSCPLSTAGSPSYVKFDNPNVGSQTF 56			

RESULT 2	STANDARD;	PRT;	869 AA.
ID CFAC_ECOLI			
AC P25733;			
DT 01-MAY-1992 (REL. 22, CREATED)			
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)			
DE CFA/I FIMBRIAL SUBUNIT C PRECURSOR (COLONISATION FACTOR ANTIGEN I			
DE SUBUNIT C).			
GN CFAC.			
OS ESCHERICHIA COLI.			
OC PLASMID NTP513.			
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;			
OC ESCHERICHIA.			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN-ENTEROTOXIGENIC;			
RX MEDLINE; 99330163.			
RA HANERS A.M., PEL H.J., WILLSHAW G.A., KUSTERS J.G.,			
RA VAN DER ZEIJST B.A.M., GAASTRA W.;			
RT *The nucleotide sequence of the first two genes of the CFA/I fimbrial			

RT operon of human enterotoxigenic Escherichia coli.;"

RL MICROB. PATHOG. 6:297-309(1989).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE; 92329981.

RA JORDI B.J.A.M., WILLSHAW G.A., VAN DER ZEIJST B.A.M., GAASTRA W.;

RT "The complete nucleotide sequence of region 1 of the CFA/I fimbrial

RT operon of human enterotoxigenic Escherichia coli.;"

RL DNA SEQ. 2:257-263(1992).

CC -!- FUNCTION: MAY SERVE AS ANCHOR FOR THE FIMBRIAE IN THE OUTER

CC MEMBRANE.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; M55661; G145510; -

DR ANTIGEN; SIGNAL; FIMBRIA; OUTER MEMBRANE; PLASMID.

FT SIGNAL 1 22 POTENTIAL.

FT CHAIN 23 869 CFA/I FIMBRIAL SUBUNIT C.

SQ SEQUENCE 869 AA; 97830 MW; 7AF76347 CRC32;

Query Match 21.1%; Score 85; DB 1; Length 869;

Best Local Similarity 32.4%; Pred. No. 6.62e-02;

Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;

Db 158 AFTOSOTINLSDGKYKRLISGNSALGITDTSY 191

QY 11 GYVQTPLESNSVRYK-LSIAGSCLPLSTAGPSY 43

RESULT 3

ID TXK\_HUMAN STANDARD; PRT; 527 AA.

AC P42681; Q14220;

DT 01-NOV-1995 (REL. 32, CREATED)

DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE TYROSINE-PROTEIN KINASE TXK (EC 2.7.1.112).

GN TXK.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BLOOD;

RX MEDLINE; 95038742.

RA HAIRE R.N., OHTA Y., LEWIS J.E., FU S.M., KROISEL P.M., LITMAN G.W.;

RT "TXK, a novel human tyrosine kinase expressed in T cells shares

RT sequence identity with Tec family kinases and maps to 4p12.;"

RL HUM. MOL. GENET. 3:897-901(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=BLOOD;

RX MEDLINE; 96197775.

RA OHTA Y., HAIRE R.N., AMEMIYA C.T., LITMAN R.T., TRAGER T., RIESS O.,

RA LITMAN G.W.;

RT "Human Txk: genomic organization, structure and contiguous physical

RT linkage with the Tec gene.;"

RL ONCOGENE 12:937-942(1996).

CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +

CC PROTEIN TYROSINE PHOSPHATE.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).

CC -!- TISSUE SPECIFICITY: EXPRESSED IN T CELLS AND SOME MYELOID CELL

CC LINES.

CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.

CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC

CC DOMAIN. BELONGS TO THE BTK SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; L27071; G684986; -

DR EMBL; G1161364; -

DR EMBL; U34379; G1161364; JOINED.

DR EMBL; U34367; G1161364; JOINED.

DR EMBL; U34368; G1161364; JOINED.

DR EMBL; U34369; G1161364; JOINED.

DR EMBL; U34370; G1161364; JOINED.

DR EMBL; U34371; G1161364; JOINED.

DR EMBL; U34372; G1161364; JOINED.

DR EMBL; U34373; G1161364; JOINED.

DR EMBL; U34374; G1161364; JOINED.

DR EMBL; U34375; G1161364; JOINED.

DR EMBL; U34376; G1161364; JOINED.

DR EMBL; U34377; G1161364; JOINED.

DR EMBL; U34378; G1161364; JOINED.

DR MIM; 600058; -

DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.

DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.

DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.

DR PROSITE; PS50001; SH2; 1.

DR PROSITE; PS50002; SH3; 1.

DR PFAM; PF00017; SH2; 1.

DR PFAM; PF00018; SH3; 1.

DR PFAM; PF00069; pkinase; 1.

DR HSP; Q06187; LAW.

KW TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;

KW SH3 DOMAIN; PHOSPHORYLATION.

FT DOMAIN 14 19 POLY-CYS.

FT DOMAIN 68 73 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).

FT DOMAIN 82 142 SH3.

FT DOMAIN 150 246 SH2.

FT DOMAIN 271 527 PROTEIN KINASE.

FT NP\_BIND 277 285 ATP (BY SIMILARITY).

FT BINDING 299 299 ATP (BY SIMILARITY).

FT ACT\_SITE 390 390 BY SIMILARITY.

FT MOD\_RES 420 420 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT CONFLICT 45 45 H -> R (IN REF. 2).

SQ SEQUENCE 527 AA; 61239 MW; 8DF019E3 CRC32;

Query Match 20.4%; Score 82; DB 1; Length 527;

Best Local Similarity 44.4%; Pred. No. 2.04e-01;

Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 242 LRYPVGLMGSCPLPATAGFSYKWEIDP 268

QY 24 VRYKSIAGSCLPLSTAGPSYVRFQDNP 50

RESULT 4

ID TXK\_MOUSE STANDARD; PRT; 527 AA.

AC P42682;

DT 01-NOV-1995 (REL. 32, CREATED)

DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE TYROSINE-PROTEIN KINASE TXK (EC 2.7.1.112) (PTK-RL-18) (RESTING

DE LYMPHOCYTE KINASE).

GN TXK OR RLK

OS MUS MUSCULUS (MOUSE).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=THYMUS;

RX MEDLINE; 96059536.

RA HAIRE R.N., LITMAN G.W.;

RT "The murine form of TXK, a novel TEC kinase expressed in thymus maps

RT to chromosome 5.,";  
 RL MAMM. GENOME 6:476-480(1995).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-FVB/N; TISSUE-THYMUS;  
 RA SOMMERS C.L., HUANG K., GRINBERG A., CHARLICK D.A., KOZAK C.A.,  
 RA LOVE P.E.;  
 RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 [3]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-C57BL/6; TISSUE=LIVER;  
 RC HIGASHITSUJI H., NONOGUCHI K., ARII S., FURUTANI M., KANEKO Y.,  
 RA NAKAYAMA H., FUJITA J.;  
 RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 [4]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE-THYMUS;  
 RC MEDLINE: 95130578  
 RA HU Q., DAVIDSON D., SCHWARTZBERG P.L., MACCHIARINI F.,  
 RA LENARDO M.J., BLUESTONE J.A., MATIS L.A.;  
 RT "Identification of Rlk, a novel protein tyrosine kinase with  
 predominant expression in the T cell lineage.";  
 RL J. BIOL. CHEM. 270:1928-1934(1995).  
 CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
 PROTEIN TYROSINE PHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN EARLY THYMOCYTES, T CELLS AND  
 MAST CELLS.  
 CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
 DOMAIN. BELONGS TO THE BTK SUBFAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL: U16145; G562125; -  
 DR EMBL: U19607; G643085; -  
 DR EMBL: D43963; G604884; -  
 DR EMBL: L35268; G623443; -  
 DR MGD: MGI:102960; TXK.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS50001; SH2; 1.  
 DR PROSITE: PS50002; SH3; 1.  
 DR PFAM: PF00017; SH2; 1.  
 DR PFAM: PF00018; SH3; 1.  
 DR PFAM: PF00069; PKINASE; 1.  
 DR HSP: O06187; 1AWW.  
 KW TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;  
 KW SH3 DOMAIN; PHOSPHORYLATION.  
 FT DOMAIN 14 20  
 FT DOMAIN 82 142  
 FT DOMAIN 150 246  
 FT DOMAIN 271 527  
 FT NP\_BIND 277 285  
 FT BINDING 299 299  
 FT ACT\_SITE 390 390  
 FT MOD\_RES 420 420  
 FT CONFLICT 3 4  
 FT CONFLICT 6 6  
 FT CONFLICT 272 272  
 FT CONFLICT 497 497  
 FT CONFLICT 497 497  
 SQ SEQUENCE 527 AA; 61108 MW; 5839DA78 CRC32;

Query Match 20.1%; Score 81; DB 1; Length 527;  
 Best Local Similarity 44.4%; Pred. No. 2.96e-01;

Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;  
 Db 242 LRVPIGLGSLPATSGFSYKWEIDP 268  
 QY 24 VRIKSIAGSCPLSTAGPSYVKFQDNP 50  
 RESULT 5  
 ID NAB3 YEAST STANDARD; PRT; 802 AA.  
 AC P38996;  
 DT 01-FEB-1995 (REL. 31, CREATED)  
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE NUCLEAR POLYADENYLATED RNA-BINDING PROTEIN NAB3.  
 GN NAB3 OR YPL190C.  
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;  
 CC SACCHAROMYCETACEAE; SACCHAROMYCES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA WILSON S.M., OBERDORF A.M., DATAR K.V., SWEDLOW J.R., PADDY M.R.,  
 RA SWANSON M.S.;  
 RL SUBMITTED (JAN-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP RIEGER M., MUELLER-AUER S., SCHAEFER M.;  
 RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNP).  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL: U05314; G476220; -  
 DR EMBL: Z73546; E246917; -  
 DR PIR: S48529; S48529.  
 DR SGD: L0001228; NAB3.  
 DR PFAM: PF00076; Rtm; 1.  
 KW RNA-BINDING; NUCLEAR PROTEIN.  
 FT DOMAIN 68 71  
 FT DOMAIN 87 93  
 FT DOMAIN 101 106  
 FT DOMAIN 108 115  
 FT DOMAIN 116 127  
 FT DOMAIN 128 137  
 FT DOMAIN 603 608  
 FT DOMAIN 644 648  
 FT DOMAIN 698 703  
 FT DOMAIN 723 728  
 FT DOMAIN 765 768  
 FT DOMAIN 769 784  
 FT DOMAIN 769 784  
 SQ SEQUENCE 802 AA; 90438 MW; FB180EDB CRC32;  
 Query Match 19.9%; Score 80; DB 1; Length 802;  
 Best Local Similarity 32.6%; Pred. No. 4.27e-01;  
 Matches 14; Conservative 10; Mismatches 16; Indels 3; Gaps 3;  
 Db 622 QCYGSGQPPPMQNSVG-RYQTSTPPPPPPQOQIPQGYRQAGP 663  
 QY 10 QGY-VQTFP-LSESNSVRYKISAGSCPLSTAGPSYVKFQDNP 50  
 RESULT 6  
 ID POLG HCV1 STANDARD; PRT; 3011 AA.  
 AC P26664;  
 DT 01-AUG-1992 (REL. 23, CREATED)  
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)



```

FT DISULFID 25 71 BY SIMILARITY.
FT DISULFID 70 79 BY SIMILARITY.
FT DISULFID 92 108 BY SIMILARITY.
FT DISULFID 107 118 BY SIMILARITY.
FT DISULFID 141 186 BY SIMILARITY.
FT DISULFID 185 194 BY SIMILARITY.
FT DISULFID 216 262 BY SIMILARITY.
FT DISULFID 261 289 BY SIMILARITY.
FT DISULFID 282 296 BY SIMILARITY.
FT DISULFID 295 307 BY SIMILARITY.
FT DISULFID 331 372 BY SIMILARITY.
FT DISULFID 371 380 BY SIMILARITY.
FT DISULFID 403 449 BY SIMILARITY.
FT DISULFID 448 458 BY SIMILARITY.
SQ SEQUENCE 472 AA; 53085 MW; 288949A5 CRC32;

Query Match 19.7%; Score 79; DB 1; Length 472;
Best Local Similarity 30.4%; Pred. No. 6.14e-01;
Matches 14; Conservative 12; Mismatches 17; Indels 3; Gaps 3;

Db 153 DFLYEYSNYGQAPLLVAYTKNY-LSVSGSCCTSANTVCFVK 197
QY 2 EKFMAEFGGYQVTPF-LSESNVRYKISAGSCPLSTAGP-SYVK 45

RESULT
ID VTDB_HUMAN STANDARD; PRT; 474 AA.
AC P02774;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE VITAMIN D-BINDING PROTEIN PRECURSOR (DBP) (GROUP-SPECIFIC COMPONENT)
DE (GC-GLOBULIN) (VDB).
GN GC.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86068030.
RA YANG F., BRUNE J.L., NAYLOR S.L., CUPPLES R.L., NABERHAUS K.H.,
RA BOWMAN B.H.;
RT "Human group-specific component (Gc) is a member of the albumin
RT family";
RL PROC. NATL. ACAD. SCI. U.S.A. 82:7994-7998(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86086396.
RA COOKE N.E., DAVID E.V.;
RT "Serum vitamin D-binding protein is a third member of the albumin and
RT alpha fetoprotein gene family.";
RL J. CLIN. INVEST. 76:2420-2424(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94092730.
RA BRAUN A., KOFER A., MORAWIETZ S., CLEVE H.;
RT "Sequence and organization of the human vitamin D-binding protein
RT gene.";
RL BIOCHIM. BIOPHYS. ACTA 1216:385-394(1993).
RN [4]
RP SEQUENCE OF 17-474.
RX MEDLINE; 86216223.
RA SCHOENTGEN F., METZ-BOUTIQUE M.-H., JOLLES J., CONSTANS J.,
RA JOLLES P.;
RT "Complete amino acid sequence of human vitamin D-binding protein
RT (group-specific component): evidence of a three-fold internal
RT homology as in serum albumin and alpha-fetoprotein.";
RL BIOCHIM. BIOPHYS. ACTA 871:189-198(1986).
RN [5]
RP SEQUENCE OF 17-31 AND 431-441.
RX MEDLINE; 79145448.
RA SVASTI J., KUROSKY A., BENNETT A., BOWMAN B.H.;
RT "Molecular basis for the three major forms of human serum vitamin D

```

```

RT binding protein (group-specific component).";
RL BIOCHEMISTRY 18:1611-1617(1979).
RN [6]
RP SEQUENCE OF 1-19 FROM N.A.
RX MEDLINE; 88005794.
RA YANG F., NABERHAUS K.H., ADRIAN G.S., GARDELLA J.M., BRISSENDEN J.E.,
RA BOWMAN B.H.;
RT "The vitamin D-binding protein gene contains conserved nucleotide
RT sequences that respond to heavy metal, adipocyte and mitotic
RT signals.";
RL GENE 54:285-290(1987).
RN [7]
RP VARIANTS GC2; GC1F AND GC1S.
RX MEDLINE; 92316509.
RA BRAUN A., BICHLMEIER R., CLEVE H.;
RT "Molecular analysis of the gene for the human vitamin-D-binding
RT protein (group-specific component): allelic differences of the common
RT genetic GC types.";
RL HUM. GENET. 89:401-406(1992).
CC -!- FUNCTION: MULTIFUNCTIONAL PROTEIN FOUND IN PLASMA, ASCITIC FLUID,
CC CEREBROSPINAL FLUID, AND URINE AND ON THE SURFACE OF MANY CELL
CC TYPES. IN PLASMA, IT CARRIES THE VITAMIN D STEROLS AND PREVENTS
CC POLYMERIZATION OF ACTIN BY BINDING ITS MONOMERS. DBP ASSOCIATES
CC WITH MEMBRANE-BOUND IMMUNOGLOBULIN ON THE SURFACE OF B-LYMPHOCYTES
CC AND WITH IGG FC RECEPTOR ON THE MEMBRANES OF T-LYMPHOCYTES.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -!- POLYMORPHISM: OVER 80 VARIANTS OF HUMAN DBP HAVE BEEN IDENTIFIED.
CC THE THREE MOST COMMON ALLELES ARE CALLED GC1F, GC1S, AND GC2. THE
CC SEQUENCE SHOWN IS THAT OF THE GC2 ALLELE.
CC -!- SIMILARITY: BELONGS TO THE ALB/AFP/VDB FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; L10641; G639896; -
EMBL; M11321; G183006; -
EMBL; X03178; G31676; -
EMBL; M12654; G181482; -
EMBL; S67480; G455970; -
EMBL; S67474; G455970; JOINED.
EMBL; S67476; G455970; JOINED.
EMBL; S67478; G455970; JOINED.
EMBL; S67479; G455970; JOINED.
EMBL; S67526; G455970; JOINED.
EMBL; M17156; G463096; -
PIR; A03237; VYHUD.
PIR; A24066; A24066.
PIR; A29096; A29096.
SWISS-2DPAGE; P02774; HUMAN.
MIN; 139200; -
PROSITE; PS00212; ALBUMIN; 1.
PFAM; PF00273; transport_prot; 1.
ALBUMIN; GLYCOPROTEIN; VITAMIN D; TRANSPORT; PLASMA; ACTIN-BINDING;
REPEAT; POLYMORPHISM; SIGNAL.
FT SIGNAL 1 16
FT CHAIN 17 474 VITAMIN D-BINDING PROTEIN.
FT REPEAT 22 199 1.
FT REPEAT 218 385 2.
FT REPEAT 404 474 3.
FT DISULFID 29 75 BY SIMILARITY.
FT DISULFID 74 83 BY SIMILARITY.
FT DISULFID 96 112 BY SIMILARITY.
FT DISULFID 111 122 BY SIMILARITY.
FT DISULFID 145 190 BY SIMILARITY.
FT DISULFID 189 198 BY SIMILARITY.
FT DISULFID 220 266 BY SIMILARITY.
FT DISULFID 265 273 BY SIMILARITY.

```



```
FT DISULFID 286 300 BY SIMILARITY.
FT DISULFID 299 311 BY SIMILARITY.
FT DISULFID 335 376 BY SIMILARITY.
FT DISULFID 375 384 BY SIMILARITY.
FT DISULFID 407 453 BY SIMILARITY.
FT DISULFID 452 462 BY SIMILARITY.
FT CARBOHYD 288 288 POTENTIAL.
FT VARIANT 432 432 D -> E (IN GC1S ALLELE).
FT VARIANT 436 436 K -> T (IN GC1F AND GC1S ALLELES).
FT CONFLICT 168 168 G -> E (IN REF. 2).
FT CONFLICT 327 327 E -> R (IN REF. 2).
SQ SEQUENCE 474 AA; 52963 MW; 56EC123D CRC32;

Query Match 19.7%; Score 79; DB 1; Length 474;
Best Local Similarity 35.7%; Pred. No. 6.14e-01;
Matches 15; Conservative 12; Mismatches 12; Indels 3; Gaps 3;

Db 157 NQFMWEYSTNGQAPLSLLVSTKSY-LSWVGSCC-TSASPT 196
QY 2 EKFMAEFGQGYVOTPF-LSESNSVRYKISAGSCPLSTAGPS 42

RESULT 9
ID YQHT_BACSU STANDARD; PRT; 353 AA.
AC P34518;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PUTATIVE PEPTIDASE IN GCVT-SPOIIIA INTERGENIC REGION (EC 3.4.-.-).
GN YQHT.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / JH642.
RA KOBAYASHI Y., MIZUNO M., MASUDA S., TAKEMARU K., HOSONO S.,
RA SATO T., TAKEUCHI M.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24B.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D84432; D1013236; -
CC DR EMBL; 299116; E1185714; -
CC DR SUBTILIST; BG11708; YQHT.
CC DR PROSITE; PS00491; PROLINE-PEPTIDASE; 1.
CC DR PFAM; PF00557; pep_M24; 1.
CC KW HYPOTHETICAL PROTEIN; HYDROLASE.
CC SQ SEQUENCE 353 AA; 38120 MW; C433A088 CRC32;

Query Match 19.4%; Score 78; DB 1; Length 353;
Best Local Similarity 29.7%; Pred. No. 8.81e-01;
Matches 11; Conservative 14; Mismatches 12; Indels 0; Gaps 0;

Db 3 LEKRNLFQGLGIDMLITSNTNRYMTGFTGSAGLA 39
QY 1 MEKFMAEFGQGYVOTPFELSESNSVRYKISAGSCPLS 37

RESULT 10
ID VTDB_RABIT STANDARD; PRT; 476 AA.
AC P53789;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE VITAMIN D-BINDING PROTEIN PRECURSOR (DBP) (GROUP-SPECIFIC COMPONENT)
```

```
DE (GC-GLOBULIN) (VDB).
GN GC OR DBP.
OS OXYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; OXYCTOLAGUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NEW ZEALAND WHITE; TISSUE=LIVER;
RX MEDLINE; 95218682.
RA OSAWA M., TSUJI T., YUKAWA N., SAITO T., TAKEICHI S.;
RT Cloning and sequence analysis of cDNA encoding rabbit vitamin
RT D-binding protein (gc globulin).;
RL BIOCHEM. MOL. BIOL. INT. 34:1003-1009(1994).
CC -1- FUNCTION: MULTIFUNCTIONAL PROTEIN FOUND IN PLASMA, ASCITIC FLUID,
CC CEREBROSPINAL FLUID, AND URINE AND ON THE SURFACE OF MANY CELL
CC TYPES. IN PLASMA, IT CARRIES THE VITAMIN D STEROLS AND PREVENTS
CC POLYMERIZATION OF ACTIN BY BINDING ITS MONOMERS. DBP ASSOCIATES
CC WITH MEMBRANE-BOUND IMMUNOGLOBULIN ON THE SURFACE OF B-LYMPHOCYTES
CC AND WITH IGG FC RECEPTOR ON THE MEMBRANES OF T-LYMPHOCYTES.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE ALB/AFP/VDB FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D29666; G603499; -
CC DR PROSITE; PS00212; ALBUMIN; 1.
CC DR PFAM; PF00273; transport_prot; 1.
CC KW ALBUMIN; GLYCOPROTEIN; VITAMIN D; TRANSPORT; PLASMA; ACTIN-BINDING;
CC REPEAT; SIGNAL.
CC FT SIGNAL 1 16 BY SIMILARITY.
CC FT CHAIN 17 476 VITAMIN D-BINDING PROTEIN.
CC FT REPEAT 22 199 1.
CC FT REPEAT 218 385 2.
CC FT REPEAT 404 476 3.
CC FT DISULFID 29 75 BY SIMILARITY.
CC FT DISULFID 74 83 BY SIMILARITY.
CC FT DISULFID 96 112 BY SIMILARITY.
CC FT DISULFID 111 122 BY SIMILARITY.
CC FT DISULFID 145 190 BY SIMILARITY.
CC FT DISULFID 189 198 BY SIMILARITY.
CC FT DISULFID 220 266 BY SIMILARITY.
CC FT DISULFID 265 273 BY SIMILARITY.
CC FT DISULFID 286 300 BY SIMILARITY.
CC FT DISULFID 299 311 BY SIMILARITY.
CC FT DISULFID 335 376 BY SIMILARITY.
CC FT DISULFID 375 384 BY SIMILARITY.
CC FT DISULFID 407 453 BY SIMILARITY.
CC FT DISULFID 452 462 BY SIMILARITY.
CC FT CARBOHYD 86 86 POTENTIAL.
CC SQ SEQUENCE 476 AA; 52911 MW; EC2DD79F CRC32;

Query Match 19.4%; Score 78; DB 1; Length 476;
Best Local Similarity 31.0%; Pred. No. 8.81e-01;
Matches 13; Conservative 15; Mismatches 11; Indels 3; Gaps 3;

Db 157 DKELYESSNYGOAPLPILVSTKSY-LSWVGTC- TSASPT 196
QY 2 EKFMAEFGQGYVOTPF-LSESNSVRYKISAGSCPLSTAGPS 42

RESULT 11
ID POLG_HCVH STANDARD; PRT; 3011 AA.
AC P27958;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
```

[illegible]



Db 146 KLSTALNQYITGRFLTKENG-RYDIVNVGGVP 177  
QY 3 KMAEFGQGYVQTPELSESNSVRYKISAGSCP 35

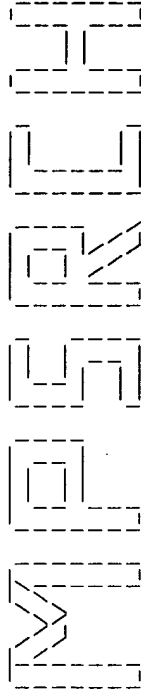
RESULT 15  
ID HEMA\_INBBO STANDARD; PRT: 575 AA.  
AC P10448;  
DT 01-MAR-1989 (REL. 10, CREATED)  
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
DE HEMAGGLUTININ PRECURSOR (FRAGMENT).  
GN HA.  
OS INFLUENZA B VIRUS (STRAIN B/BONN/43).  
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; ORTHOMYXOVIRIDAE;  
OC INFLUENZA VIRUS A AND B GROUP.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA RICHIE L.R., AIR G.M.;  
RL SUBMITTED (JAN-1989) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS  
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X13550; G758202; -  
DR PIR; S01882;  
DR PFAM; PF00509; Hemagglutinin; 1.  
DR HSP; P03437; SHMG.  
KW ENVELOPE PROTEIN; HEMAGGLUTININ; GLYCOPROTEIN; SIGNAL.  
FT NON\_TER 1 1  
FT SIGNAL <1 8  
FT CHAIN 9 351 HA1 CHAIN.  
FT CHAIN 353 575 HA2 CHAIN.  
FT CARBOHYD 33 33 POTENTIAL.  
FT CARBOHYD 67 67 POTENTIAL.  
FT CARBOHYD 153 153 POTENTIAL.  
FT CARBOHYD 156 156 POTENTIAL.  
FT CARBOHYD 171 171 POTENTIAL.  
FT CARBOHYD 238 238 POTENTIAL.  
FT CARBOHYD 309 309 POTENTIAL.  
FT CARBOHYD 338 338 POTENTIAL.  
FT CARBOHYD 497 497 POTENTIAL.  
FT CARBOHYD 523 523 POTENTIAL.  
FT CARBOHYD 536 536 POTENTIAL.  
FT CARBOHYD 568 568 POTENTIAL.  
SQ SEQUENCE 575 AA; 62318 MW; E31DFE5 CRC32;

Query Match 18.7%; Score 75; DB 1; Length 575;  
Best Local Similarity 50.0%; Pred. No. 2.54e+00;  
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 143 YK1TSGSCPNTVNGSGF 160  
QY 26 YK1SIAGSCPLSTAGPSY 43

Search completed: Wed Sep 1 16:09:53 1999  
Job time : 10 secs.

\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

MParch_pp      protein - protein database search, using Smith-Waterman algorithm
Run on:        Wed Sep 1 16:10:10 1999;   MasPar time 7.61 Seconds
              401.470 Million cell updates/sec
Tabular output not generated.

```

```
>PCT-US99-13024-2
Title:
Description: (1-56) from PCTUS9913024.pep (4 of 12)
Perfect Score: 402
Sequence: 1 MEKMAEFGGGYVQPFLE.....STAGPS
```

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl9  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 34.913; Variance 62.861; scale 0.555

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query %		Length	DB	ID	Description	Pred. No.
	Score	Match					
1	81	20.1	3011	14	Q03463	GENOME POLYPROTEIN.	1.08e+00
2	80	19.9	275	14	Q68488	POLYPROTEIN (FRAGMENT)	1.50e+00
3	80	19.9	275	14	Q68467	POLYPROTEIN (FRAGMENT)	1.50e+00
4	80	19.9	275	14	Q68469	POLYPROTEIN (FRAGMENT)	1.50e+00
5	80	19.9	357	5	Q01993	VC27A7L.1 PROTEIN.	1.50e+00
6	80	19.9	802	3	Q07034	RNA BINDING PROTEIN.	1.50e+00
7	80	19.9	2435	14	Q81756	POLYPROTEIN (FRAGMENT)	1.50e+00
8	79	19.7	542	5	Q17456	SIMILAR TO GLUTAMATE D	2.10e+00
9	79	19.7	860	1	Q59003	860AA LONG HYPOTHETICA	2.10e+00
10	78	19.4	70	2	P75688	FROM BASES 311709 TO 3	2.91e+00
11	78	19.4	3011	14	Q36610	POLYPROTEIN.	2.91e+00
12	78	19.4	3011	14	Q36608	POLYPROTEIN.	2.91e+00
13	78	19.4	3011	14	Q36609	POLYPROTEIN.	2.91e+00
14	78	19.4	3011	14	Q36579	POLYPROTEIN.	2.91e+00
15	77	19.2	418	2	Q56631	LECITHINASE.	4.04e+00
16	77	19.2	470	2	Q87325	LECITHINASE.	4.04e+00
17	77	19.2	566	2	Q84354	HYPOTHETICAL 63.5 KD P	4.04e+00
18	77	19.2	967	2	Q54123	PEPB.	4.04e+00
19	77	19.2	1464	11	Q63728	N-METHYL-D-ASPARTATE R	4.04e+00
20	77	19.2	1464	11	Q08948	N-METHYL-D-ASPARTATE R	4.04e+00

77	19.2	1464	4	Q12879	N-METHYL-D-ASPARTATE R	4.04e+00
21	22.9	2219	5	Q23388	SMILK1067.2 PROTEIN.	4.04e+00
77	19.2	76	18.9	535	22	5.58e+00
77	18.9	535	5	P91315	23	5.58e+00
76	18.9	550	5	Q02490	24	5.58e+00
76	18.9	821	2	O51735	25	5.58e+00
76	18.9	821	5	O51735	26	5.58e+00
76	18.9	1142	4	Q14497	27	5.58e+00
76	18.9	1142	5	Q14497	28	5.58e+00
76	18.9	2660	5	Q19785	29	5.58e+00
76	18.9	364	3	Q74624	30	5.70e+00
78	18.7	364	3	Q74624	31	5.70e+00
74	18.4	978	3	P06074	32	1.06e+01
74	18.4	978	5	P06074	33	1.06e+01
74	18.4	121	3	Q07070	34	1.06e+01
73	18.2	92	14	Q72154	35	1.45e+01
32	32	42	4	Q72154	36	1.45e+01
73	18.2	482	4	Q43411	37	1.45e+01
33	33	586	9	Q38005	38	1.45e+01
73	18.2	586	9	Q38005	39	1.45e+01
73	18.2	964	2	Q53784	40	1.45e+01
73	18.2	1693	14	O81876	41	1.45e+01
73	18.2	2301	10	O80784	42	1.45e+01
36	36	37	72	17.9	37	1.45e+01
72	17.9	347	14	Q82650	43	1.98e+01
38	38	347	14	Q82658	44	1.98e+01
72	17.9	347	14	Q82658	45	1.98e+01
39	39	362	14	Q67383	46	1.98e+01
72	17.9	362	14	Q67383	47	1.98e+01
41	41	362	14	Q67381	48	1.98e+01
72	17.9	379	14	Q67363	49	1.98e+01
42	42	379	14	Q67363	50	1.98e+01
72	17.9	379	14	Q67366	51	1.98e+01
43	43	379	14	Q67366	52	1.98e+01
72	17.9	447	14	O89483	53	1.98e+01
72	17.9	447	14	O89483	54	1.98e+01
72	17.9	447	14	O89496	55	1.98e+01

## ALIGNMENTS

```

RESULT 1
ID Q03463 PRELIMINARY; PRT: 3011 AA.
AC Q03463;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DT GENE POLYPROTEIN.
OS HEPATITIS C VIRUS (ISOLATE HC-J1) (HCV).
OC VIRUSES; SRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE.
CC HEPATITIS C-LIKE VIRUSES.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE: 93117120.
RX OKAMOTO H., KANAI N., MISHIRO S.;
RA OKAMOTO H., OKADA S., SUGIYAMA Y., YOTSUMOTO S., TANAKA T.,
RA YOSHITAZA H., TSUDA F., MIYAKAWA Y., MAYUMI M.;
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RN JPN. J. EXP. MED. 60:167-177(1990).
DR EMBL: D10749; G221587; -
DR PFAM: PF00998; HCV_RGRP; 1.
DR PFAM: PF01001; HCV_NS4b; 1.
DR PFAM: PF01006; HCV_NS4a; 1.
KW POLYPROTEIN.
SQ SEQUENCE 3011 AA; 327114 MW; 399EDBFA CRC32;

Query Match 20.1%; Score 81; DB 14; Length 3011;
Best Local Similarity 40.0%; Pred. No. 1.08e+00;
Matches 18; Conservative 10; Mismatches 13; Indels 4; Gaps

Db 2116 VPSPEFTTLDGVRLLH-REAPDCKPDLLEEVGFVGLHDYVGSQ 2159
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 13 VQTP-ELSESNSRYKISIAAGSC-PLSTAGPSY-VKFDQNPVGSQ 54

RESULT 2
ID Q68468 PRELIMINARY; PRT: 275 AA.
AC Q68468;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

```









[illegible]





Search completed: Wed Sep 1 16:10:34 1999  
Job time : 24 secs.

\*\*\*\*\*

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

MPsrch_pp  protein - protein database search, using Smith-Waterman algorithm
Run on:      Wed Sep 1 16:14:29 1999;  MasPar time 5.88 Seconds
                281.992 Million cell updates/sec
Tabular output not generated.

```

```
>PCT-US99-13024-2
Title:
Description:
Perfect Score:
Sequence:
Scoring table:

(1-78) from PCTUS9913024.pep (5 of 12)
551
1 MEXFMAEFGQGVYQTFPSE.....GLHLRFVDFSTGALVDKSY 78

PAM 150
Gap 11
```

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: a-geneseq35

1: part1 2: part2 3: part3 4: part4 5: part5 6: part6 7: part7  
8: part8 9: part9 10: part10 11: part11 12: part12 13: part13  
14: part14 15: part15 16: part16 17: part17 18: part18  
19: part19 20: part20 21: part21 22: part22 23: part23  
24: part24 25: part25 26: part26 27: part27 28: part28  
29: part29 30: part30 31: part31 32: part32 33: part33  
34: part34 35: part35 36: part36 37: part37 38: part38  
39: part39

Statistics: Mean 26.730; Variance 108.537; scale 0.246

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description	Pred. No.
		Match	Length					
1	300	54.4	65	19	R97371		Phage T4-ORF4-gene pr	1.22e-19
2	96	17.4	1693	35	W76368		Hepatitis E virus hol	1.66e+00
3	87	15.8	623	35	W5773		Amino acid sequence o	9.05e+00
4	86	15.6	1693	10	R51264		HEV strain protein en	1.09e+01
5	86	15.6	1693	36	W80196		Protein encoded by OR	1.09e+01
6	86	15.6	1693	34	R91813		Hepatitis E virus str	1.09e+01
7	86	15.6	1693	34	W1209		Protein encoded by OR	1.09e+01
8	86	15.6	1693	37	W81519		Hepatitis E virus (HE	1.09e+01
9	85	15.4	633	35	W75774		Amino acid sequence o	1.31e+01
10	84	15.2	1693	3	R14618		Protein encoded by OR	1.57e+01
11	82	14.9	633	35	W75775		Amino acid sequence o	2.27e+01
12	81	14.7	3011	12	R66995		Hepatitis C virus gen	2.72e+01
13	81	14.7	3011	8	R40120		HCV genomic amino aci	2.72e+01
14	80	14.5	170	1	R92033		Sequence encoded in t	3.26e+01
15	80	14.5	170	1	R90150		Sequence of hepatitis	3.26e+01
16	80	14.5	411	16	R90934		HCV NS5 domain antiq	3.26e+01

17	80	14.5	499	38	W67010	HCV non-structural pr	3.26e+01
18	80	14.5	509	20	W08378	<i>Massica napus</i> micros	3.26e+01
19	80	14.5	509	2	R06519	Microspore-specific c	3.26e+01
20	80	14.5	516	6	R33575	HCV KKS-NS5E recombin	3.26e+01
21	80	14.5	516	4	R21566	HCV KKS-NS5E - pHCV-4	3.26e+01
22	80	14.5	516	6	R33441	HCV NS5 encoded by pH	3.26e+01
23	80	14.5	516	7	R33633	HCV KKS-NS5E fusion p	3.26e+01
24	80	14.5	516	6	R33595	HCV KKS-NS5E recombin	3.26e+01
25	80	14.5	798	7	R33630	HCV KKS-NS5 EF fusion	3.26e+01
26	80	14.5	1766	1	P92041	Sequence encoded in t	3.26e+01
27	80	14.5	1786	1	P90158	Protein sequence of h	3.26e+01
28	80	14.5	2261	1	P90164	Peptide encoded by co	3.26e+01
29	80	14.5	2301	1	P92047	Sequence encoded in t	3.26e+01
30	80	14.5	2435	5	R25335	HCV polypeptide 1.	3.26e+01
31	80	14.5	2436	5	R28582	HCV amino acid sequen	3.26e+01
32	80	14.5	2436	1	P92050	Sequence encoded in t	3.26e+01
33	80	14.5	2462	1	P90288	Peptide encoded by co	3.26e+01
34	80	14.5	2772	2	R08123	Hepatitis C virus pol	3.26e+01
35	80	14.5	2816	7	R34009	HCV-1 polyprotein.	3.26e+01
36	80	14.5	2894	5	R24440	Composite HCV Hc-J1/C	3.26e+01
37	80	14.5	2894	13	R70230	Composite hepatitis C	3.26e+01
38	80	14.5	2955	2	R08124	Hepatitis C virus put	3.26e+01
39	80	14.5	3011	16	R90931	Hepatitis C virus pol	3.26e+01
40	80	14.5	3011	4	R21519	Compiled HCV sequence	3.26e+01
41	80	14.5	3011	26	W34480	HCV polyprotein.	3.26e+01
42	80	14.5	3011	28	W40038	HCV polyprotein.	3.26e+01
43	80	14.5	3011	6	R31621	Hepatitis C virus (HC	3.26e+01
44	79	14.3	458	23	W10561	Vitamin D3-binding pr	3.91e+01
45	79	14.3	458	11	R56975	Vitamin D-binding b bin	3.91e+01

## ALIGNMENTS

```

RESULT 1
ID R97371 standard; Protein; 65 AA.
AC R97371;
DE 07-JAN-1997 (first entry)
DT Phase T4 ORFX gene product_gp34.
KW Phase T4; tail fibre protein; nanotechnology; nano-structure;
KW filter; molecular sieve.
OS Bacteriophage T4.
PN WO9611947-A1.
PD 25-APR-1996.
PF 13-OCT-1995; U13023.
PR 13-OCT-1994; US-322760.
PA (GOLD/) GOLDBERG E B.
PI Goldberg EB;
DR WPI: 96-221942/22.
DR N-PSDB; T23053.
DR New proteins derived from T4 phage tail fibre proteins - that can
PT self assemble into nano-structure(s), useful as filters etc, also
PT corresponding DNA
PS Claim 7; Fig 7; 83pp; English.
CC A protein (R97371) of unspecified function is the product of open
CC reading frame x of the tail fibre protein gene region (see also
CC T29053) of phage T4. This gene region also includes open reading
CC frames for tail fibre proteins (see also R97370 and R97372-74).
CC Tail fibre proteins (native or modified) can be produced in large
CC quantities in microbial cells and used as bulding blocks of strong
CC stable nanostructures.
SQ Sequence 65 AA;

Query Match 54.4%; Score 300; DB 19; Length 65;
Best Local Similarity 82.1%; Pred. No. 1.22e-19;
Matches 46; Conservative 3; Mismatches 7; Indels 0; Gaps

Db 1 mekfmaelwtricpnaalssesvrvkisiagscplstagspsvxfqdnpgvsqtf 56
|||||: ::::::::::::::::::::::::::::::::::::::
QY 1 MERKMAEFGGVYQTFPLESSESVRVKISIAGSCPLSTAGPSVYVFQDNPGVSQTF 56

RESULT 2
ID W76368 standard; protein; 1693 AA.

```



[illegible]

05-JUL-1990; US-505888.  
07-JUN-1995; US-475807.  
PA (GENE-) GENELABS TECHNOLOGIES INC.  
PI Bradley DW, Fry KE, Krawczynski KZ, Reyes GR, Tam A,  
PI Yarbrough PO;  
DR WPI: 98-582599/49.  
DR N-PSDB: V66321.  
PT Hepatitis E virus proteins - useful for diagnosis or vaccine  
production the virus  
PS Claim 22; Columns 57-66; 47pp; English.  
CC W80196-98 are encoded by the genome of the Burma strain of  
CC enterically transmitted non-A non-B hepatitis virus (ET-NANB)  
CC (hepatitis E virus (HEV)). The specification describes an isolated  
CC protein which is specifically immunoreactive with antibodies present  
CC in individuals infected with HEV and encoded by a sequence contained  
CC in an open reading frame (ORF) of an HEV genome. The genome has a  
CC sequence that is more than 70% identical to the ORF1 sequence from  
CC Burma HEV isolate. The protein is used as a vaccine and a diagnostic  
CC probe for ET-NANB.  
CC Sequence 1693 AA;

Query Match 15.6%; Score 86; DB 36; Length 1693;  
Best Local Similarity 25.4%; Pred. No. 1.09e+01;  
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 frtsfdgavletngperhnlsfdaqstmaagpfsltyaaasaglevryvaagldhrav 631  
QY 8 FCGGYVTQPFLESNSRVRYKISAGSCPLSTAGPSYVKFDQNPVGSGT-F-SAGLHLR-V 64  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |  
Db 632 fapgvsp 638  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |  
QY 65 FDPSTGA 71

RESULT 6  
ID R91813 standard; Protein; 1693 AA.  
AC R91813;  
DT 26-NOV-1996 (first entry)  
DE Hepatitis E virus strain SAR-55 ORF-1.  
KW Hepatitis E virus; HEV; SAR-55 strain; enteric transmission;  
KW structural region; antigen; detection; antibody; vaccine;  
KW immunisation; infection.  
OS Hepatitis E virus.  
FH Key Location/Qualifiers  
FT misc\_difference 1238 /note= "corresponding codon CAG"  
FT misc\_difference 1455..1693  
FT /note= "10 bp nucleic acid sequence TGGNTTYGA  
has to be inserted between nucleotides  
4390..4391 of T27394 before these amino  
acid residues can be decoded"

W09610580-A2.  
PD 11-APR-1996.  
PF 03-OCT-1995; U13102.  
PR 03-OCT-1994; US-316765.  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PI Emerson SU, Purcell RH, Tsarev SA;  
DR WPI: 96-209320/21.  
DR N-PSDB: T27394.

PT Isolated and purified hepatitis E virus strain SAR-55 DNA - encodes  
PT antigenic protein useful in diagnosis, prophylaxis and treatment of  
PT hepatitis E virus infection  
PS Disclosure: pages 9-13; 121pp; English.  
CC The present sequence is the protein prod. of ORF-1 from the  
CC hepatitis E virus (HEV) strain SAR-55, which was implicated in an  
CC enterically transmitted non-A, non-B hepatitis in Pakistan. The  
CC protein encoded by the structural region of the virus (i.e. ORF-2),  
CC which is capable of forming HEV like particles, is useful for the  
CC detection of HEV antibodies (pref. IgG or IgM) in blood, plasma,  
CC sera, cerebrospinal fluid, tissue, urine or pleural fluid. The  
CC protein, and anti-HEV antibodies generated using the protein, can  
CC also be used in vaccines for immunising an animal against HEV  
CC infection. The protein is identified as a band of greater than

05 kD following SDS-PAGE of cell lysates of insect cells infected  
with a HEV ORF-2 contg. baculovirus, i.e. the claimed recombinant  
expression vectors pPIC9-1779, -1780 and -1781.

Query Match 15.6%; Score 86; DB 19; Length 1693;  
Best Local Similarity 25.4%; Pred. No. 1.09e+01;  
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 frtsfdgavletngperhnlsfdaqstmaagpfsltyaaasaglevryvaagldhrav 631  
QY 8 FCGGYVTQPFLESNSRVRYKISAGSCPLSTAGPSYVKFDQNPVGSGT-F-SAGLHLR-V 64  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |  
Db 632 fapgvsp 638  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |  
QY 65 FDPSTGA 71

RESULT 7  
ID W71209 standard; Protein; 1693 AA.  
AC W71209;  
DT 30-OCT-1998 (first entry)  
DE Protein encoded by ORF 1 of the Burmese isolate of ET-NANB.  
KW Enterically transmitted nonA/nonB hepatitis virus; identification;  
KW HEV; ET-NANB; detection; vaccine.  
OS Hepatitis virus.  
FH Key Location/Qualifiers  
FT Misc\_difference 154 /note= "not specified"  
FT Misc\_difference 1514 /note= "not specified"  
FT Misc\_difference 1552 /note= "not specified"  
FN US5789559-A.  
PD 04-AUG-1998.  
PF 25-JUN-1994; 279823.  
PR 05-APR-1991; US-681078.  
PR 17-JUN-1988; US-208997.  
PR 11-APR-1989; US-336672.  
PR 19-JUN-1989; US-367486.  
PR 13-OCT-1989; US-420921.  
PR 05-APR-1990; US-505888.  
PR 25-JUL-1994; US-279823.  
PA (GENE-) GENELABS TECHNOLOGIES INC.  
PI Bradley DW, Fry KE, Krawczynski KZ, Reyes GR, Tam A,  
PI Yarbrough PO;  
DR WPI: 98-446186/38.  
DR N-PSDB: V54729.

PT Hepatitis E virus DNA - useful for e.g. virus detection and viral  
PT protein production  
PS Disclosure: Columns 55-64; 45pp; English.  
CC W71209-11 represent the proteins encoded by the open reading frames  
CC (ORFs) of the DNA sequence of the Burmese isolate of an enterically  
CC transmitted nonA/nonB viral hepatitis agent (ET-NANB). The nucleic  
CC acid sequence may be used for identifying and sequencing the entire  
CC viral agent (also referred to as HEV), detecting ET-NANB in  
CC infected samples, e.g. by specific amplification of virus-derived DNA  
CC sequences and for producing recombinant viral proteins for use in  
CC vaccines.

Query Match 15.6%; Score 86; DB 34; Length 1693;  
Best Local Similarity 25.4%; Pred. No. 1.09e+01;  
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 frtsfdgavletngperhnlsfdaqstmaagpfsltyaaasaglevryvaagldhrav 631  
QY 8 FCGGYVTQPFLESNSRVRYKISAGSCPLSTAGPSYVKFDQNPVGSGT-F-SAGLHLR-V 64  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |  
Db 632 fapgvsp 638  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |  
QY 65 FDPSTGA 71





Search completed: Wed Sep 1 16:15:06 1999  
Job time : 37 secs.

(TM)

Distribution rights by Oxford Molecular Ltd

500.410 Million cell updates/sec

## SUMMARIES

Result No.	Query			DB	ID	Description	Pred. No.
	Score	Match	Length				
1	92	16.7	299	2	S61248	hypothetical protein	7.69e-02
2	89	16.2	566	2	G71525	probable OMP [leader	2.14e-01
3	86	15.6	1693	1	MNWHE	genome polyprotein -	5.84e-01
4	85	15.4	331	2	A61046	ecdysone-induced memb	8.12e-01
5	85	15.4	869	2	C56617	cfaC protein precursor	8.12e-01
6	84	15.2	622	2	S17402	parasporal crystal pr	1.13e+00
7	83	15.1	281	2	S38913	hypothetical protein	1.56e+00
8	83	15.1	611	2	D70928	hypothetical protein	1.56e+00
9	82	14.9	81	2	I38375	tyrosine kinase - hum	2.15e+00
10	82	14.9	449	2	S55092	hypothetical protein	2.15e+00
11	82	14.9	527	2	I84483	tyrosine kinase - hum	2.15e+00
12	81	14.7	216	2	G70447	flagellar L-ring prot	2.96e+00
13	81	14.7	527	2	A55631	protein-tyrosine kina	2.96e+00
14	81	14.7	527	2	I49133	Txk - mouse	2.96e+00
15	81	14.7	3011	2	S40770	polyprotein precursor	2.96e+00
16	80	14.5	171	2	B38162	hypothetical protein	4.05e+00
17	80	14.5	572	2	S14200	GRESAG protein - Tryp	4.05e+00
18	80	14.5	802	2	S48529	NAB3 protein - yeast	4.05e+00
19	80	14.5	954	2	S46105	glucan 1,4-alpha-gluc	4.05e+00
20	80	14.5	1659	2	JC4956	vitellogenin precursor	4.05e+00
21	80	14.5	3011	1	GNWVC3	genome polyprotein -	4.05e+00
22	79	14.3	370	2	S69178	hypothetical protein	5.55e+00
23	79	14.3	472	1	A35327	vitamin D-binding pro	5.55e+00



```
#cross-references MUID:99000809
#accession G71525
##status preliminary
##molecule_type DNA
##residues 1-566 ##label ARN
##cross-references GB:AE001308; GB:AE001273; NID:93328766; PID:g3328771
##experimental_source serotype D, strain UW-3/Cx
GENETICS
#gene
#accession CT350
#length 566 #molecular-weight 63507 #checksum 4960
SUMMARY
Query Match 16.2%; Score 89; DB 2; Length 566;
Best Local Similarity 26.2%; Pred. No. 2.14e-01;
Matches 17; Conservative 25; Mismatches 18; Indels 5; Gaps 5;
Db 63 IAEYLOQSFLESDTVIR-KSAIIGA-GLSGSEA-LELLSEAIETQDLYEQL-L-ILNA 117
QY 8 FCGYVQTFPLESSESVRYKISAGSCPLSTAGPSYVKFQDNPVGSQTFSAGLHLRVFDP 67
Db 118 ATSQL 122
QY 68 STGAL 72
RESULT 3
ENTRY MNWHE #type complete
TITLE genome polyprotein - hepatitis E virus (strain Burma)
CONTAINS RNA-directed RNA polymerase (EC 2.7.7.48)
ORGANISM #formal_name hepatitis E virus
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
29-May-1998
ACCESSIONS A40778; A48547
REFERENCE A40778
#authors Tam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, G.R.
#journal Virology (1991) 185:120-131
#title Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome.
#cross-references MUID:92024067
#accession A40778
##molecule_type genomic RNA
##residues 1-1693 ##label TAM
##cross-references GB:W73218; NID:9330023; PID:g330024
REFERENCE A48547
#authors Fry, K.E.; Tam, A.W.; Smith, M.M.; Kim, J.P.; Luk, K.C.; Young, L.M.; Piatk, M.; Feldman, R.A.; Yun, K.Y.; Purdy, M.A.; McCaustland, K.A.; Bradley, D.W.; Reyes, G.R.
#journal Virus Genes (1992) 6:173-185
#title Hepatitis E virus (HEV): strain variation in the nonstructural gene region encoding consensus motifs for an RNA-dependent RNA polymerase and an ATP/GTP binding site.
#cross-references MUID:92271462
#accession A48547
##molecule_type genomic RNA
##residues 967-1693 ##label FRY
##cross-references GB:M32400; NID:9330021; PID:g330022
##note sequence extracted from NCBI backbone (NCBIN:104572, NCBI:P:104573)
CLASSIFICATION #superfamily hepatitis E virus nonstructural protein
KEYWORDS ATP; nonstructural protein; nucleotidyltransferase
SUMMARY #length 1693 #molecular-weight 185191 #checksum 6520
Query Match 15.6%; Score 86; DB 1; Length 1693;
Best Local Similarity 25.4%; Pred. No. 5.84e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;
Db 572 FRTSFVDSGAVLTNGPERNLSDASQSTMAAGPSLTYAASAGLEVYVVAAGLDHRAV 631
QY 8 FCGYVQTFPLESSESVRYKISAGSCPLSTAGPSYVKFQDNPVGSQT-F-SAGLHLR-V 64
Db 632 FAPGVSP 638
QY 65 FDPSTGA 71
```

```
RESULT 4
ENTRY A61046 #type complete
TITLE ecdysone-induced membrane protein IMP-E3 - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Feb-1997
ACCESSIONS A61046
REFERENCE A61046
#authors Moore, J.T.; Fristrom, D.; Hammonds, A.S.; Fristrom, J.W.
#journal Dev. Genet. (1990) 11:299-309
#title Characterization of IMP-E3, a gene active during imaginal disc morphogenesis in Drosophila melanogaster.
#accession A61046
##status preliminary
##molecule_type mRNA
##residues 1-331 ##label MOO
GENETICS
#gene FlyBase:ImpE3
##cross-references FlyBase:Fbgn0001255
KEYWORDS membrane protein
SUMMARY #length 331 #molecular-weight 36583 #checksum 8221
Query Match 15.4%; Score 85; DB 2; Length 331;
Best Local Similarity 31.9%; Pred. No. 8.12e-01;
Matches 15; Conservative 12; Mismatches 18; Indels 2; Gaps 2;
Db 190 LNFRLRLYDNTGRAAFGESAMDRMSTASIAGKKRVPTKP-YVDF 235
QY 1 MEKFNAEFGQGVQTFPLESSESVRYKI-SIAGSCPLSTAGPSYVKF 46
RESULT 5
ENTRY C56617 #type complete
TITLE cfac protein precursor - Escherichia coli plasmid NTP113
ORGANISM #formal_name Escherichia coli
DATE 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 20-Mar-1998
ACCESSIONS C56617
REFERENCE A56617
#authors Jordi, B.J.; Willshaw, G.A.; van der Zeijst, B.A.; Gastra, W.
#journal DNA Seq. (1992) 2:257-263
#title The complete nucleotide sequence of region 1 of the CFA/I fimbrial operon of human enterotoxigenic Escherichia coli.
#cross-references MUID:92329981
#accession C56617
##status preliminary
##molecule_type DNA
##residues 1-869 ##label JOR
##cross-references GB:M55861; NID:g145507; PID:g145510
##experimental_source enterotoxigenic strain, CFA/I-St plasmid NTP113
##note sequence extracted from NCBI backbone (NCBIN:108960, NCBI:P:108971)
GENETICS
#gene cfac
#genome plasmid
SUMMARY #length 869 #molecular-weight 97830 #checksum 9755
Query Match 15.4%; Score 85; DB 2; Length 869;
Best Local Similarity 32.4%; Pred. No. 8.12e-01;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;
Db 158 AFIQSQTNLSDSGYKKRLISGNSALGITDTSY 191
QY 11 GYVQTFPLESSESVRYK-ISIAGSCPLSTAGPSY 43
RESULT 6
ENTRY SI7402 #type complete
TITLE parasporal crystal protein cryIIc - Bacillus thuringiensis
```

```

ALTERNATE_NAMES
ORGANISM    delta-endotoxin
DATE        22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
09-Sep-1997
ACCESSIONS  S17402
REFERENCE    S17400
#authors     Wu, D.; Cao, X.L.; Bai, Y.Y.; Aronson, A.I.
#journal     FEMS Microbiol. Lett. (1991) 81:31-36
#title       Sequence of an operon containing a novel delta-endotoxin gene
            from Bacillus thuringiensis.
#accession   S17402
##status     preliminary
##molecule_type DNA
##residues   1-622 ##label WUD
##cross-references EMBL:X57252; NID:g40283; PID:g40286
GENETICS
#genome      plasmid
#keywords     delta-endotoxin
SUMMARY      #length 622 #molecular-weight 69729 #checksum 8205
Query Match  15.2%; Score 84; DB 2; Length 622;
Best Local Similarity 18.2%; Pred. No. 1.13e+00;
Matches 12; Conservative 26; Mismatches 26; Indels 2; Gaps 2;
Db 501 FISEYKQGDRLRELNPTARYTLRNGNSYNYLRVSSIGSTIRVINGRYTANV 560
QY 12 YVQTPFLSESNVRYKISAGS-CPLSTAGPSYVKF-QDNPGVSGTFSAGLHLRVFDPST 69
Db 561 NTTTTN 566
QY 70 GALVDS 75

RESULT 7
ENTRY   S38913 #type complete
TITLE   hypothetical protein 1 - phage phi-C31
ORGANISM #formal_name phage phi-C31
DATE     20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
09-Sep-1997
ACCESSIONS S38913
REFERENCE    S38912
#authors     Hartley, N.M.; Murphy, G.O.; Bruton, C.J.; Chater, K.F.
#submission  submitted to the EMBL Data Library, November 1993
#accession   S38913
##status     preliminary
##molecule_type DNA
##residues   1-281 ##label HAR
##cross-references EMBL:X76288; NID:g432610; PID:g579071
GENETICS
#start_codon GTG
SUMMARY      #length 281 #molecular-weight 31680 #checksum 2790
Query Match  15.1%; Score 83; DB 2; Length 281;
Best Local Similarity 24.1%; Pred. No. 1.56e+00;
Matches 14; Conservative 17; Mismatches 24; Indels 3; Gaps 3;
Db 194 AYMNADFIIDGNREPMPEFDGAVALHVTDETWA-FKPVTGPDVFAQFLHLRQTFD 250
QY 11 GYVQTPFLSESNVRYKIS-ITAGSCPLSTAGPSYVKFQDNPGVSGTFSAGLHLR-VFD 66

RESULT 8
ENTRY   D70928 #type complete
TITLE   hypothetical protein Rv2913c - Mycobacterium tuberculosis
            (strain H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE     17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS D70928
REFERENCE    A70500
#authors     Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
            C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry

```

```

III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
Nature (1998) 393:537-544
Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession D70928
##status     preliminary; nucleic acid sequence not shown;
            translation not shown
##molecule_type DNA
##residues_type 1-611 ##label COL
##cross-references GB:274024; GB:AL123456; NID:g3250700; PID:e1301028;
            PID:g3250707
##experimental_source strain H37Rv
GENETICS
#gene         Rv2913c
SUMMARY      #length 611 #molecular-weight 67204 #checksum 9971
Query Match  15.1%; Score 83; DB 2; Length 611;
Best Local Similarity 43.5%; Pred. No. 1.56e+00;
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
Db 319 VRFQHLVPFFELYSDGIDLVPFE 341
QY 44 VKFDQNPVGSQTFSAGLHLRVFD 66

RESULT 9
ENTRY   I38375 #type fragment
TITLE   tyrosine kinase - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE     29-May-1998 #sequence_revision 29-May-1998 #text_change
10-Jul-1998
ACCESSIONS I38375
REFERENCE    I38372
#authors     Haire, R.N.; Ohta, Y.; Lewis, J.E.; Fu, S.M.; Kroisel, P.;
            Litman, G.W.
#journal     Hum. Mol. Genet. (1994) 3:897-901
#title       TXK, a novel human tyrosine kinase expressed in T cells
            shares sequence identity with Tec family kinases and maps
            to 4p12.
#cross-references MUID:95038742
#accession   I38375
##status     preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues_type 1-81 ##label RES
##cross-references EMBL:U07794; NID:g508219; PID:g508224
GENETICS
#gene         GDB:TXK
#cross-references GDB:377329; OMIM:600058
#map_position 4p12-4p12
#introns      25/1
CLASSIFICATION #superfamily protein-tyrosine kinase tec; pleckstrin repeat
            homology; protein kinase homology; SH2 homology; SH3
            homology
FEATURE
#domain        #domain protein kinase homology (fragment) #label KIN
SUMMARY      #length 81 #checksum 7751
Query Match  14.9%; Score 82; DB 2; Length 81;
Best Local Similarity 44.4%; Pred. No. 2.15e+00;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Db 5 LRYVGLMGSLCPATAGFSYEKWEIDP 31
QY 24 VRYKISAGSCPLSTAGPSYVKFQDNP 50

```

```
RESULT 10
ENTRY
TITLE
ALTERNATE_NAMES
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#submissions
#accession
#residues
#molecule_type
#cross-references
#experimental_source
GENETICS
#map_position
CLASSIFICATION
SUMMARY
Query Match
Best Local Similarity
Matches
Db
QY
RESULT 11
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references
#accession
#residues
#molecule_type
#cross-references
#map_position
CLASSIFICATION
FEATURE
89-137
150-246
269-527
SUMMARY
Query Match
Best Local Similarity
Matches
Db
QY
RESULT 12
ENTRY
TITLE
ALTERNATE_NAMES
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references
#accession
#residues
#molecule_type
#cross-references
#experimental_source
GENETICS
SUMMARY
Query Match
Best Local Similarity
Matches
Db
QY
RESULT 13
ENTRY
TITLE
ALTERNATE_NAMES
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references
#accession
#status
#molecule_type
#residues
#cross-references
CLASSIFICATION
KEYWORDS
FEATURE
89-137
150-246
269-527
277-285
SUMMARY
Query Match
```

Search completed: Wed Sep 1 16:14:10 1999  
Job time : 19 secs.

Result No.	Query			DB	ID	Description	Pred. No.
	Score	Match	Length				
1	300	54.4	65	1	Y15A_BPT4	HYPOTHETICAL 7.3 KD PR	6.06e-47
2	100	18.1	1693	1	ENH8_PICVGL	NON-STRUCTURAL POLYPRO	9.43e-04
3	86	15.6	457	1	POLN_HEVBU	LATE EMBRYOGENESIS ABU	1.92e-01
4	86	15.6	1693	1	POLN_HEVBU	NON-STRUCTURAL POLYPRO	1.92e-01
5	86	15.6	1693	1	POLN_HEVPA	NON-STRUCTURAL POLYPRO	1.92e-01
6	85	15.4	703	1	Y41I_RH1SN	HYPOTHETICAL 76.2 KD P	2.79e-01
7	85	15.4	869	1	CFAC_ECOLI	CFA/I FIMBRIAL SUBUNIT	2.75e-01
8	83	15.1	611	1	YX45_MYCTU	HYPOTHETICAL 67.2 KD P	5.61e-01
9	82	14.9	449	1	Y460_YEAST	HYPOTHETICAL 51.4 KD P	7.96e-01
10	82	14.9	527	1	KD2_HUMAN	TYROSINE-PROTEIN KINAS	7.96e-01
11	81	14.7	527	1	TKX_MOUSE	TYROSINE-PROTEIN KINAS	1.13e+00
12	80	14.5	171	1	YCB2_PSEDE	HYPOTHETICAL 19.0 KD P	1.59e+00
13	80	14.5	572	1	CY42_TRYBB	RECEPTOR-TYPE ADENYLAT	1.59e+00
14	80	14.5	802	1	NAB3_YEAST	NUCLEAR POLYADENYLATED	1.59e+00
15	80	14.5	954	1	YB79_YEAST	PUTATIVE FAMILY 31 GLU	1.59e+00
16	80	14.5	1097	1	IMB3_HUMAN	IMPORTIN BETA-3 SUBUNIT	1.59e+00
17	80	14.5	1659	1	VIT1_ONCMY	VITELLOGENIN PRECURSOR	1.59e+00
18	80	14.5	3011	1	POLG_HCVI	GENOME POLYPROTEIN [CO	1.59e+00
19	79	14.3	472	1	VTD8_MOUSE	VITAMIN D-BINDING PROT	2.24e+00
20	79	14.3	474	1	VTD8_HUMAN	VITAMIN D-BINDING PROT	2.24e+00
21	79	14.3	633	1	CR72_BACTK	70 KD CRYSTAL PROTEIN	2.24e+00
22	79	14.3	935	1	YJ31_YEAST	HYPOTHETICAL 108.4 KD	2.24e+00
23	78	14.2	353	1	YQHT_BACSU	PUTATIVE PEPTIDASE IN	3.14e+00



```

CC -----
DR EMBL; M80381; G329998; -
KW POLYPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; HELICASE;
KW ATP-BINDING.
FT NP_BIND 975 982 ATP (POTENTIAL).
SQ SEQUENCE 1693 AA; 185149 MW; FBCA2483 CRC32;

Query Match 15.6%; Score 86; DB 1; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.92e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 FRTSFDGAVLTNGPERNLNLSFDSQSTMAAGPESLTVAASAGLEVYVVAAGLDHRAV 631
QY 8 FCGYVQTPFLSESNSVRYKISAGSCPLSTAGSPSYVKFQDNPVGSQT-F-SAGLHLR-V 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 632 FAGVSP 638
QY 65 FDPSTGA 71
| : : :
| : : :

RESULT 6
ID Y411 RHISN STANDARD; PRT; 703 AA.
AC P55492;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 76.2 KD PROTEIN Y411.
GN Y411.
OS RHIZOBIUM SP. (STRAIN NGR234).
OG PLASMID SYN FNGR234A.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
CC RHIZOBIACEAE; RHIZOBIUM.
CC [1]
CC SEQUENCE FROM N.A.
RX MEDLINE; 97305956.
RA FREIBERG C.A., FELLAY R., BAIRUCH A., BROUGHTON W.J., ROSENTHAL A.,
RA PERRET X.;
RT Molecular basis of symbiosis between Rhizobium and legumes.*;
RL NATURE 387:394-401(1997).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: NONE OBVIOUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000078; G2182446; -
KW HYPOTHETICAL PROTEIN; PLASMID; TRANSMEMBRANE.
FT TRANSHEM 23 43 POTENTIAL.
FT TRANSHEM 69 89 POTENTIAL.
FT TRANSHEM 143 163 POTENTIAL.
FT TRANSHEM 250 270 POTENTIAL.
FT TRANSHEM 357 377 POTENTIAL.
FT TRANSHEM 432 452 POTENTIAL.
FT TRANSHEM 644 664 POTENTIAL.
SQ SEQUENCE 703 AA; 76183 MW; A2BA53CE CRC32;

Query Match 15.4%; Score 85; DB 1; Length 703;
Best Local Similarity 27.0%; Pred. No. 2.75e-01;
Matches 17; Conservative 17; Mismatches 28; Indels 1; Gaps 1;

Db 224 SPFADTDSGTSTAATSLFGAGAGATRPPEWLAFAAALASGAPFGLSPRVATATATSOYA 283
QY 15 TPFL-SESNSVRYKISAGSCPLSTAGSPSYVKFQDNPVGSQTF-SAGLHLRVPDSTGALV 73
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 284 KAK 286
QY 74 DSK 76
| :
| :

```

```

RESULT 7
ID CFAC_ECOLI STANDARD; PRT; 869 AA.
AC P25733;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE CFA/I FIMBRIAL SUBUNIT C PRECURSOR (COLONISATION FACTOR ANTIGEN I
DE SUBUNIT C).
GN CFAC.
OS ESCHERICHIA COLI.
OG PLASMID NTP513.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
CC [1]
CC SEQUENCE FROM N.A.
RX STRAIN-ENTEROTOXIGENIC;
RX MEDLINE; 89330163.
RA HAMERS A.M., PEL H.J., WILLSHAW G.A., KUSTERS J.G.,
RA VAN DER ZEIJST B.A.M., GAASTRA W.;
RT "The nucleotide sequence of the first two genes of the CFA/I fimbrial
RT operon of human enterotoxigenic Escherichia coli.";
RL MICROB. PATHOG. 6:297-309(1989).
CC [2]
CC SEQUENCE FROM N.A.
RX MEDLINE; 92329981.
RA JORDI B.J.A.M., WILLSHAW G.A., VAN DER ZEIJST B.A.M., GAASTRA W.;
RT "The complete nucleotide sequence of region 1 of the CFA/I fimbrial
RT operon of human enterotoxigenic Escherichia coli.";
RL DNA SEQ. 2:257-263(1992).
CC -!- FUNCTION: MAY SERVE AS ANCHOR FOR THE FIMBRIAE IN THE OUTER
CC MEMBRANE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M55661; G145510; -
KW ANTIGEN; SIGNAL; FIMBRIA; OUTER MEMBRANE; PLASMID.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 869 CFA/I FIMBRIAL SUBUNIT C.
SQ SEQUENCE 869 AA; 97830 MW; 7AF76347 CRC32;

Query Match 15.4%; Score 85; DB 1; Length 869;
Best Local Similarity 32.4%; Pred. No. 2.75e-01;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;

Db 158 AFIOQTINLSDSGYKRLISGNSALGITDTSY 191
QY 11 GYVQTPFLSESNSVRYK-ISIAGSCPLSTAGPSY 43
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
ID YX45_MYCTU STANDARD; PRT; 611 AA.
AC Q10830;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 67.2 KD PROTEIN CY274.45C.
GN MYC274.45C OR MYC338.01C.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIADAE;
OC ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
CC [1]
CC SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RA CONNOR R., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: BELONGS TO THE N-ACYL-D-AMINO-ACID DEACYLASE FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z74697; E249958; -
DR EMBL: Z74024; E248895; -
KW HYPOTHETICAL PROTEIN; HYDROLASE.
SQ SEQUENCE 611 AA; 67205 MW; 9DF140EE CRC32;

Query Match 15.1%; Score 83; DB 1; Length 611;
Best Local Similarity 43.5%; Pred. No. 5.61e-01;
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Db 319 VRFQHLVPFPELYSGDIDLPVFE 341
QY 44 VKFDNPVGSOTFSAGLHLRVFD 66

RESULT 9
ID YM60-YEAST STANDARD; PRT; 449 AA.
AC Q03649;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 51.4 KD PROTEIN IN RARI-SCJ1 INTERGENIC REGION.
GN YMR210W OR YMR261.04.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA DEDMAN K., BROWN D., BOWMAN S., BARRELL B.G., RAJANDREAM M.A.,
RA WALSH S.V.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: BELONGS TO THE UPF0017 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z49809; G854462; -
DR PROSITE: PS01133; UPF0017; 1.
DR PFAM: PF00561; abhydrolase; 1.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 449 AA; 51437 MW; FE884F32 CRC32;

Query Match 14.9%; Score 82; DB 1; Length 449;
Best Local Similarity 27.1%; Pred. No. 7.96e-01;
Matches 16; Conservative 18; Mismatches 22; Indels 3; Gaps 3;

Db 227 YMGFGSLGASIMTYLGEESDRTKIECAISVSNPFDLYNSAYF-INSTPMGSRFYSPAL 284
QY 4 FMAEFGQG-VYQTFPLSE-SNSVRYKISAGSCPLSTAGSPSYKVFQDNPVGSQTFSSAGL 60

RESULT 10
ID TXK_HUMAN STANDARD; PRT; 527 AA.
AC P42681; Q14220;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE TXK (EC 2.7.1.112).
KW TXK.

```

```

OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BLOOD;
RX MEDLINE: 95038742.
RA HAIRE R.N., OHTA Y., LEWIS J.E., FU S.M., KROISEL P.M., LITMAN G.W.;
RT "TXK, a novel human tyrosine kinase expressed in T cells shares
RL sequence identity with Tec family kinases and maps to 4p12.";
RL HUM. MOL. GENET. 3:897-901(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BLOOD;
RX MEDLINE: 96197775.
RA OHTA Y., HAIRE R.N., ANEMIYA C.T., LITMAN R.T., TRAGER T., RIESS O.,
RA LITMAN G.W.;
RT "Human Txk: genomic organization, structure and contiguous physical
RL linkage with the Tec gene.";
RL ONCOGENE 12:937-942(1996).
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN T CELLS AND SOME MYELOID CELL
CC LINES.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE BTK SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L27071; G684986; -
DR EMBL: U34379; G1161364; -
DR EMBL: U34367; G1161364; JOINED.
DR EMBL: U34368; G1161364; JOINED.
DR EMBL: U34369; G1161364; JOINED.
DR EMBL: U34370; G1161364; JOINED.
DR EMBL: U34371; G1161364; JOINED.
DR EMBL: U34372; G1161364; JOINED.
DR EMBL: U34373; G1161364; JOINED.
DR EMBL: U34374; G1161364; JOINED.
DR EMBL: U34375; G1161364; JOINED.
DR EMBL: U34376; G1161364; JOINED.
DR EMBL: U34377; G1161364; JOINED.
DR EMBL: U34378; G1161364; JOINED.
DR MIM: 600058; -
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS50001; SH2; 1.
DR PROSITE: PS50002; SH3; 1.
DR PFAM: PF00017; SH2; 1.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00069; pkinase; 1.
DR HSP: Q06187; LAWW.
KW TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;
KW SH3 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 14 19 POLY-CYS.
FT DOMAIN 68 73 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 82 142 SH3.
FT DOMAIN 150 246 SH2.
FT DOMAIN 271 527 PROTEIN KINASE.
FT NP_BIND 277 285 ATP (BY SIMILARITY).
FT BINDING 299 299 ATP (BY SIMILARITY).
FT ACT_SITE 390 390 BY SIMILARITY.
FT MOD_RES 420 420 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

```



```
FT CONFLICT 45 45 H -> R (IN REF. 2).
SQ SEQUENCE 527 AA: 61239 MW: 8DF019E3 CRC32;

Query Match
Best Local Similarity 44.4%; Pred. No. 7.96e-01;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 242 LRYVGLMGSCPLPATAGSYEKEWIDP 268
   :||::|||:|||||:|:|:|
QY 24 VRYKISAGCPLSTAGPSYVAFQDNP 50

RESULT 11
ID TXK_MOUSE STANDARD; PRT: 527 AA.
AC P42682;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE TXK (EC 2.7.1.112) (PTK-RL-18) (RESTING
DE LYMPHOCYTE KINASE).
GN TXK OR RLK.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE-THYMUS;
RX MEDLINE: 96059536.
RA HAIRE R.N., LITMAN G.W.;
RT "The murine form of TXK, a novel TEC kinase expressed in thymus maps
RT to chromosome 5."
RL MAMM. GENOME 6:476-480(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-FVB/N; TISSUE-THYMUS;
RA SOMMERS C.L., HUANG K., GRINBERG A., CHARLICK D.A., KOZAK C.A.,
RA LOVE P.E.;
RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6; TISSUE=LIVER;
RA HIGASHITSUJI H., NONOGUCHI K., ARII S., FURUTANI M., KANEKO Y.,
RA NAKAYAMA H., FUJITA J.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=THYMUS;
RX MEDLINE: 95130578.
RA HU Q., DAVIDSON D., SCHWARTZBERG P.L., MACCHIARINI F.,
RA LENARDO M.J., BLUESTONE J.A., MATIS L.A.;
RT "Identification of Rlk, a novel protein tyrosine kinase with
RT predominant expression in the T cell lineage."
RL J. BIOL. CHEM. 270:1928-1934(1995).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN EARLY THYMOCYTES, T CELLS AND
CC MAST CELLS.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE BTK SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U16145; G562125; -
EMBL; U19607; G643065; -
```

```
DR EMBL; D43963; G604884; -
DR EMBL; L35268; G623443; -
DR MGD; MGI:102960; TXK.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSP; Q06187; LAWW.
DR TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;
KW SH3 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 14 20 POLY-CYS.
FT DOMAIN 82 142 SH3.
FT DOMAIN 150 246 SH2.
FT NP_BIND 271 285 PROTEIN KINASE.
FT BINDING 299 299 ATP (BY SIMILARITY).
FT ACT_SITE 390 390 ATP (BY SIMILARITY).
FT MOD_RES 420 420 BY SIMILARITY.
FT CONFLICT 3 4 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 6 6 LS -> SF (IN REF. 3).
FT CONFLICT 272 272 Y -> D (IN REF. 3).
FT CONFLICT 497 497 A -> T (IN REF. 3).
FT CONFLICT 497 497 R -> S (IN REF. 3).
SQ SEQUENCE 527 AA; 61108 MW; 5B39DA78 CRC32;

Query Match 14.7%; Score 81; DB 1; Length 527;
Best Local Similarity 44.4%; Pred. No. 1.13e-00;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 242 LRYVGLMGSCPLPATAGSYEKEWIDP 268
   :||::|||:|||||:|:|:|
QY 24 VRYKISAGCPLSTAGPSYVAFQDNP 50

RESULT 12
ID YCB2_PSEDE STANDARD; PRT: 171 AA.
AC P29944;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 19.0 KD PROTEIN IN COBS 5' REGION (ORF2).
OS PSEUDOMONAS DENITRIFICANS.
OC BACTERIA; PROTEOBACTERIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92011364.
RA CAMERON B., GUILHOT C., BLANCHE F., CAUCHOIS L., ROUYEZ M.-C.,
RA RIGAUD S., LEVY-SCHIL S., CROUZET J.;
RT "Genetic and sequence analyses of a Pseudomonas denitrificans DNA
RT fragment containing two cob genes."
RJ J. BACTERIOL. 173:6058-6065(1991).
CC -!- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M62869; G151167; -
PIR; B38162; B38162.
DR PROSITE; PS00636; DNAJ_1; FALSE_NEG.
DR PROSITE; PS50076; DNAJ_2; 1.
DR PFAM; PF00226; DnaJ; 1.
DR HSP; P08622; 1XBL.
DR HYPOTHETICAL PROTEIN.
SQ SEQUENCE 171 AA; 18973 MW; 5DD2D1D5 CRC32;
```

```
Query Match 14.5%: Score 80; DB 1; Length 171;
Best Local Similarity 37.5%: Pred. No. 1.59e+00;
Matches 24; Conservative 11; Mismatches 25; Indels 4; Gaps 4;

Db 24 EYNGKYNFFSGLDSEVARQKEAITHRPTWTVGVN-KNAKNQPTQSQRSGSAGAAQAR 82
QY 7 EFGQGYVQTPLFSESNVRY-KISAGSCPLSTAGPSYKRVQDNVPVGSQT-F-SAGLHLR 63

Db 83 MRDP 86
QY 64 VFDP 67

RESULT 13
ID CV42.TRYBB STANDARD; PRT; 572 AA.
AC Q9396;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE RECEPTOR-TYPE ADENYLATE CYCLASE GRESAG 4.2 (EC 4.6.1.1) (ATP
DE PYROPHOSPHATE-LYASE) (ADENYL CYCLASE) (FRAGMENT).
GN GRESAG 4.2.
OS TRYPANOSOMA BRUCEI BRUCEI.
OC EUKARYOTA; EUGLENOZOA; KINETOPLASTIDA; TRYPANOSOMATIDAE; TRYPANOSOMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EATRO 1125;
RX MEDLINE: 91218809.
RA ALEXANDRE S., PAINDAVIONE P., TEBABI P., PAYS A., HALLEUX S.,
RA STEINERT M., PAYS E.;
RT "Differential expression of a family of putative adenylate/guanylate
RT cyclase genes in Trypanosoma brucei.";
RL MOL. BIOCHEM. PARASITOL. 43:279-288(1990).
CC -1- FUNCTION: COULD ACT AS A RECEPTOR FOR A UNKNOWN LIGAND.
CC -1- CATALYTIC ACTIVITY: ATP -> 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC -1- SIMILARITY: BELONGS TO CLASS-3 OF ADENYL CYCLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; X52120; E30923;
DR PFAM; PF00211; guanylate_cyc; 1.
DR HSP; Q02846; JAWL.
DR KW LYASE; CAMP SYNTHESIS; TRANSMEMBRANE; RECEPTOR; GLYCOPROTEIN.
FT NON_TER 1 1 EXTRACELLULAR (POTENTIAL).
FT DOMAIN <1 225 POTENTIAL.
FT TRANSMEM 226 251 POTENTIAL.
FT DOMAIN 252 572 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 10 10 POTENTIAL.
FT CARBOHYD 77 77 POTENTIAL.
FT CARBOHYD 152 152 POTENTIAL.
SQ SEQUENCE 572 AA; 63683 MW; CE3202BF CRC32;

Query Match 14.5%: Score 80; DB 1; Length 572;
Best Local Similarity 30.0%: Pred. No. 1.59e+00;
Matches 15; Conservative 16; Mismatches 17; Indels 2; Gaps 2;

Db 442 LSTAERSQFDVTLGGVPLRGVSEPVYQNLNAPVGRSF-AELRLDRVLD 490
QY 18 LSESNVRYKISAGSCPLSTAGPSYKRVQDNVPVGSOTFSAGLHL-RVFD 66

RESULT 14
ID NAB3.YEAST STANDARD; PRT; 802 AA.
AC P38996;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)

Query Match 14.5%: Score 80; DB 1; Length 802;
Best Local Similarity 32.6%: Pred. No. 1.59e+00;
Matches 14; Conservative 10; Mismatches 16; Indels 3; Gaps 3;

Db 622 QGYGSOPIPMNOSYG-RYOTSIPPPPPQOQIPOGYGRYQAGP 663
QY 10 QGY-VQTPF-LSESNVRYKISAGSCPLSTAGPSYKRVQDNVP 50

RESULT 15
ID YB79.YEAST STANDARD; PRT; 954 AA.
AC P38138;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PUTATIVE FAMILY 31 GLUCOSIDASE IN PCS60-ABD1 INTERGENIC REGION
DE (EC 3.2.1.-).
GN YBR229C OR YBR1526.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RA DUBOIS E., EL BAKKOURY M., GLANSDORFF N., MESSENGUY F., PIERARD A.,
RA SCHERENS B., VIERENDELS F.;
RA SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN
```

